

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
*Center for Advanced Study
in the Behavioral Sciences*

LEONARD S. KOGAN
*Community Service Society
of New York*

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: BARBARA CUMMINGS; FRANCES H. CLARK; SADIE J. DOYLE

VOLUME 54
1957



PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1957 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 54

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

	PAGE
ARGYLE, M. Social Pressure in Public and Private Situations.....	172
AUSTRIH H. See SCODEL, A.	
BAGBY, J. W. A Cross-Cultural Study of Perceptual Predominance in Binocular Rivalry.....	331
BAKER, BELA O., & BLOCK, J. Accuracy of Interpersonal Prediction as a Function of Judge and Object Characteristics.....	37
BARRY, H., III. Relationships Between Child Training and the Pictorial Arts.....	380
BASOWITZ, H., & KORCHIN, S. J. Age Differences in the Perception of Closure.....	93
BASOWITZ, H. See KORCHIN, S. J.	
BASS, B. M. Reply to Messick and Jackson's Comments on Authoritarianism or Acquiescence.....	426
BEHAR, I., & RIOPELLE, A. J. The Effects of Reserpine on the Acquisition of the Conditioned Avoidance Response.....	160
BELL, R. Q. See SCHAEFER, E. S.	
BENSEN, S. B. See SEIDMAN, D.	
BENTON, A. L., & BLACKBURN, H. L. Practice Effects in Reaction-Time Tasks in Brain-Injured Patients.....	109
BERKOWITZ, L. Liking for the Group and the Perceived Merit of the Group's Behavior.....	353
BERRY, J. L., & MARTIN, B. GSR Reactivity as a Function of Anxiety, Instructions, and Sex.....	9
BLACKBURN, H. L. See BENTON, A. L.	
BLOCH, BARBARA L. See MAUSNER, B.	
BLOCK, J. A Comparison Between Ipsative and Normative Ratings of Personality.....	50
BLOCK, J. See BAKER, BELA O.	
BLOCK, J. Studies in the Phenomenology of Emotions.....	358
BRACKBILL, YVONNE, & LITTLE, K. B. Factors Determining the Guessing of Meanings of Foreign Words.....	312
BRIM, O. G., JR., & HOFF, D. B. Individual and Situational Differences in Desire for Certainty.....	225
BROEN, W. E., JR. A Factor-Analytic Study of Religious Attitudes.....	176
BROWN, R. W., LEITER, R. A., & HILDUM, D. C. Metaphors from Music Criticism.....	347
BURWEN, L. S., & CAMPBELL, D. T. The Generality of Attitudes Toward Authority and Nonauthority Figures.....	24
CAMPBELL, D. T. See BURWEN, L. S.	
CAMPBELL, D. T. See CHAPMAN, L. J.	
CATTELL, R. B., STICE, G. F., & KRISTY, N. F. A First Approximation to Nature-Nurture Ratios for Eleven Primary Personality Factors in Objective Tests.....	143
CERVIN, V. Relationship of Ascendant-Submissive Behavior in Dyadic Groups of Human Subjects to Their Emotional Responsiveness.....	241
CHAPMAN, L. J., & CAMPBELL, D. T. Response Set in the F Scale.....	129
CHAPMAN, L. J., & TAYLOR, JANET A. Breadth of Deviate Concepts Used by Schizophrenics.....	118
CHURCH, R. M. Transmission of Learned Behavior Between Rats.....	163
CLARK, J. W. See KAMIN, L. J.	
COHEN, A. S. See STOTLAND, E.	
CRASILNECK, H. B., & MICHAEL, CARMEN M. Performance on the Bender Under Hypnotic Age Regression.....	319
CROW, W. J., & HAMMOND, K. R. The Generality of Accuracy and Response Sets in Interpersonal Perception.....	384
CULBERTSON, FRANCES M. Modification of an Emotionally Held Attitude Through Role Playing.....	230
DAVIS, R. H., & HARRINGTON, R. W. The Effect of Stimulus Class on the Problem-Solving Behavior of Schizophrenics and Normals.....	126
DITTES, J. E. Extinction During Psychotherapy or GSR Accompanying "Embarrassing" Statements.....	187
EHRLICH, DANUTA, GUTTMAN, I., SCHONBACH, P., & MILLS, J. Postdecision Exposure to Relevant Information.....	98
ERIKSEN, C. W. See KUETHE, J. L.	
ERRATA.....	428
FESHBACH, S., & SINGER, R. The Effects of Personal and Shared Threats Upon Social Prejudice.....	411
FINDLAY, D. C., & MCGUIRE, C. Social Status and Abstract Behavior.....	135
FINE, B. J. Conclusion-Drawing, Communicator Creditability, and Anxiety as Factors in Opinion Change.....	369
FREEMAN, J. T., & MAHER, H. A Note on "Discrimination Reaction Time as a Function of Anxiety and Intelligence".....	134
FRENCH, J. R. P., JR., & ZAJONC, R. B. An Experimental Study of Cross-Cultural Norm Conflict.....	218
GLASSMAN, S. M. See HARDING, G. F.	
GORDON, J. E. Leading and Following Psychotherapeutic Techniques with Hypnotically Induced Repression and Hostility.....	405

GREENBANK, R. K. The Wolf.....	PAGE 41
GRISSELL, J. L. See ROSENBAUM, G.	
GRUNES, WILLA F. Looking at Occupations.....	8
GUTTMAN, I. See EHRLICH, DANUTA.	
GYNTHNER, RUTH A. The Effects of Anxiety and of Situational Stress on Communicative Efficiency.....	274
HAMMOND, K. R. See CROW, W. J.	
HARDING, G. F., GLASSMAN, S. M., & HELZ, W. C. Maturation and the Spiral Aftereffect.....	276
HARDY, K. R. Determinants of Conformity and Attitude Change.....	289
HARRINGTON, R. W. See DAVIS, R. H.	
HEARN, G. Leadership and the Spatial Factor in Small Groups.....	269
HELZ, W. C. See HARDING, G. F.	
HILDUM, D. C. See BROWN, R. W.	
HOEHN, A. J. See SALTZ, E.	
HOFF, D. B. See BRIM, O. G., JR.	
HOVLAND, C. I., & PRITZKER, H. A. Extent of Opinion Change as a Function of Amount of Change Advocated.....	257
HOWES, D. On the Relation Between the Probability of a Word as an Association and In General Linguistic Usage.....	75
JACKSON, D. N. See MESSICK, S.; SOLLEY, C. M.	
JACKSON, D. N., & MESSICK, S. J. A Note on "Ethnocentrism" and Acquiescent Response Sets.....	132
JACKSON, D. N., MESSICK, S. J., & SOLLEY, C. M. How "Rigid" is the "Authoritarian?".....	137
JENSEN, A. R. Authoritarian Attitudes and Personality Maladjustment.....	303
JOHNSON, L. C. Generality of Speed and Confidence of Judgment.....	264
KAESS, W. A. See WITRYOL, S. L.	
KAMENETZKY, J., & SCHMIDT, H. Effects of Personal and Impersonal Refutation of Audience Counter- arguments on Attitude Change.....	200
KAMIN, L. J., & CLARK, J. W. The Taylor Scale and Reaction Time.....	262
KELMAN, H. C., & PARLOFF, M. B. Interrelations Among Three Criteria of Improvement in Group Therapy: Comfort, Effectiveness, and Self-Awareness.....	281
KESSEN, W. The Role of Experience in Judging Children's Photographs.....	375
KLEINER, R. See PEPITONE, A.	
KLEINMAN, M. L. Psychogenic Deafness and Perceptual Defense.....	335
KORCHIN, S. J. See BASOWITZ, H.	
KORCHIN, S. J., & BASOWITZ, H. Age Differences in Verbal Learning.....	64
KORCHIN, S. J., & LEVINE, S. Anxiety and Verbal Learning.....	234
KRISTY, N. F. See CATTELL, R. B.	
KUETHE, J. L., & ERIKSEN, C. W. Personality, Anxiety, and Muscle Tension as Determinants of Response Stereotypy.....	400
LEITER, R. A. See BROWN, R. W.	
LEVINE, S. See KORCHIN, S. J.	
LITTLE, K. B. See BRACKBILL, YVONNE.	
MACKAVEY, W. R. See ROSENBAUM, G.	
MCGAUGHRAN, L. S., & MORAN, L. J. Differences Between Schizophrenic and Brain-Damaged Groups in Conceptual Aspects of Object Sorting.....	44
MCGUIRE, C. See FINDLAY, D. C.	
MAHER, B. A. Personality, Problem Solving, and the Einstellung Effect.....	70
MAHER, H. See FREEMAN, J. T.	
MARTIN, B. See BERRY, J. L.	
MASON, D. J. Judgments of Leadership Based Upon Physiognomic Cues.....	273
MATARAZZO, J. D. See SASLOW, G.	
MATARAZZO, RUTH G. See SASLOW, G.	
MAUSNER, B., & BLOCH, BARBARA L. A Study of the Additivity of Variables Affecting Social Inter- action.....	250
MEELAND, T. See SEIDMAN, D.	
MESSICK, S., & JACKSON, D. N. Authoritarianism or Acquiescence in Bass's Data.....	424
MESSICK, S. J. See JACKSON, D. N.	
MESSICK, S. J. See SOLLEY, C. M.	
MICHAEL, CARMEN M. See CRASILNECK, H. B.	
MILLER, I. See SEIDMAN, D.	
MILLS, J. See EHRLICH, DANUTA.	
MORAN, L. J. See MCGAUGHRAN, L. S.	
NUTHMANN, ANNE M. Conditioning of a Response Class on a Personality Test.....	19
PAPAJOHNS, J. See SPIVAK, M.	
PARLOFF, M. B. See KELMAN, H. C.	
PEPITONE, A., & KLEINER, R. The Effects of Threat and Frustration on Group Cohesiveness.....	192

PERKINS, J. C. See PODELL, L.	
PHARES, E. J. Expectancy Changes in Skill and Chance Situations.....	339
PHILLIPS, JEANNE S. See SASLOW, G.	
PICKREL, E. W. Levels of Perceptual Organization and Cognition: Conflicting Evidence.....	422
PODELL, L., & PERKINS, J. C. A Guttman Scale for Sexual Experience—A Methodological Note.....	420
PRITZKER, H. A. See HOVLAND, C. I.	
RABINOWITZ, W. Anality, Agression, and Acquiescence.....	140
RAPHIELSON, A. C. The Relationships Among Imaginative, Direct Verbal, and Physiological Measures of Anxiety in an Achievement Situation.....	13
RIOPELLE, A. J. See BEHAR, I.	
ROSENBAUM, G., MACKAVEY, W. R., & GRISSELL, J. L. Effects of Biological and Social Motivation on Schizophrenic Reaction Time.....	364
ROTHSCHILD, G. H. See SHAW, M. E.	
SALTZ, E., & HOHEN, A. J. A Test of the Taylor-Spence Theory of Anxiety.....	114
SARASON, I. G. Effect of Anxiety and Two Kinds of Motivating Instructions on Verbal Learning.....	166
SASLOW, G., MATARAZZO, J. D., PHILLIPS, JEANNE S., & MATARAZZO, RUTH G. Test-Retest Stability of Interaction Patterns During Interviews Conducted One Week Apart.....	295
SCHAEFER, E. S., & BELL, R. Q. Patterns of Attitudes Toward Child Rearing and the Family.....	391
SCHMIDT, H. See KAMENETZKY, J.	
SCHONBACH, P. See EHRLICH, DANUTA.	
SCODEL, A., & AUSTRIN, H. The Perception of Jewish Photographs by Non-Jews and Jews.....	278
SEIDMAN, D., BENSEN, S. B., MILLER, I., & MEELAND, T. Influence of a Partner on Tolerance for a Self- Administered Electric Shock.....	210
SHAW, M. E., ROTHSCCHILD, G. H., & STRICKLAND, J. F. Decision Processes in Communication Nets.....	323
SHOR, R. E. Effect of Preinformation Upon Human Characteristics Attributed to Animated Geo- metric Figures.....	124
SINGER, R. See FESHBACH, S.	
SMITH, H. P. The Effects of Intercultural Experience—A Follow-up Investigation.....	266
SOLLEY, C. M., JACKSON, D. N., & MESSICK, S. J. Guessing Behavior and Autism.....	32
SOLLEY, C. M. See JACKSON, D. N.	
SPENCE, D. P. A New Look at Vigilance and Defense.....	103
SPIVAK, M., & PAPAJOHNS, J. The Effect of the Schedule of Reinforcement on Operant Conditioning of a Verbal Response in the Autokinetic Situation.....	213
STECKLER, G. A. Authoritarian Ideology in Negro College Students.....	396
STICE, G. F. See CATTELL, R. B.	
STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A. S., & ZANDER, A. The Effects of Group Expectations and Self-Esteem Upon Self-Evaluation.....	55
STRICKLAND, J. F. See SHAW, M. E.	
TAYLOR, JANET A. See CHAPMAN, L. J.	
THOMAS, E. See STOTLAND, E.	
THORLEY, S. See STOTLAND, E.	
TROW, D. B. Autonomy and Job Satisfaction in Task-Oriented Groups.....	204
VEROFF, J. Development and Validation of a Projective Measure of Power Motivation.....	1
WEISS, W. Opinion Congruence with a Negative Source on One Issue as a Factor Influencing Agree- ment on Another Issue.....	180
WITRYOL, S. L., & KAES, W. A. Sex Differences in Social Memory Tasks.....	343
ZAJONC, R. B. See FRENCH, J. R. P., JR.	
ZANDER, A. See STOTLAND, E.	

DEVELOPMENT AND VALIDATION OF A PROJECTIVE MEASURE OF POWER MOTIVATION¹

JOSEPH VEROFF

Survey Research Center, University of Michigan

THE purpose of this study was to develop a valid measure of power motivation. McClelland and his associates have shown that scoring stories written in response to pictures can be used as a basis for measures of strength of the achievement (8) and the affiliation (3, 11) motives. These investigators have shown that with an increase in experimentally aroused achievement or affiliation motivation, there was in the stories an increase in the frequency of certain kinds of imagery related to these motives. Furthermore, the coding categories, which can be used to reflect these shifts in imagery can be successfully employed as a basis for measuring individual differences in achievement and affiliation motivations. Subjects who have high scores on these measures are significantly different from subjects who have low scores in behaviors which are meaningfully related to the strength of the motives.

Following the same general procedures of these studies, the experimenter anticipated that a parallel measure of another motive—the power motive—could be developed.

What is meant by power motivation? Numerous social theorists have used a concept of power. In some cases, the concept is relegated to the part it plays in the analysis of social structure. In other cases, social power is a dimension to be considered in the analysis of interpersonal relationships. In the latter sense, the power motive will be considered that disposition directing behavior toward satisfactions contingent upon the control of the means of influencing another person(s). In the phenomenal sphere of the power-motivated individual, he considers himself the "gate-keeper" to certain decision-making of others. The means of control can be anything at all

that can be used to manipulate another person. Overt dominance strivings can be considered one kind of control execution. The definition of the power motive offered here, however, is meant to include more than dominance. The importance of this kind of social motivation has been elaborated by Adler (1) and Sullivan (12) and has been brought to bear upon some of the recent research in group dynamics (6).

Before one can make these predictions about the relationship of power motivation to behavior, a means of measuring the strength of the motive has to be established. The design of this research to develop such a measuring instrument falls into two steps.

First, written stories are obtained in response to pictures from two groups, one for whom it can be assumed that power motivation is aroused, and the other for whom it can be assumed the concern about power is not specifically aroused. A scoring system isolating the kind of imagery presumably reflecting power motivation can then be applied to these two sets of stories. The stories from the aroused group should contain more imagery in the categories scored for power motivation than those from the nonaroused group. Differences in imagery obtained between the two groups provide an empirical clarification of how power motivation is expressed in thought processes.

Second, the expected differences in imagery produced in these two experimental groups are used as the basis for a measure of individual differences in power motivation. Scores so derived should predict individual differences in other behaviors related to satisfying power motivation.

EXPERIMENTAL DESIGN AND HYPOTHESES

Step 1: Group Differences in Power Motivation

For the *arousal condition* the experimenter was able to take advantage of a natural situation: the election of student leaders at the University of Michigan. The candidates must petition for their candidacy. Once having declared candidacy, they have a month's time

¹ The research reported in this paper was submitted to the faculty of the University of Michigan in partial fulfillment of the requirements for the Ph.D. degree. The author would like to acknowledge the assistance of Dr. Daniel Katz in directing this work. The project was carried out while the author was a Research Training Fellow of the Social Science Research Council. He is indebted to the Council for its financial aid.

for campaigning before the actual election. After two days of balloting, the candidates congregate at the polls to see what the balloting has decided. At least two hours elapse after they have congregated before any official returns are announced. It was therefore decided to employ this waiting period for the study. All candidates who had at least one other opponent were previously asked by mail to cooperate in a research project designed to get at important social variables. The only information given was that the experimenter was interested in them as campus leaders, and was interested in the effects of social situations upon imagination. Because the candidates were presumably extremely concerned about the control of the means of influence—being elected to office—the situation was thought to meet the requirements for arousal of the motive as it was defined.

Since not all the candidates participated, volunteering bias may have been introduced. Such bias, however, would seem irrelevant to the problem under investigation. All 34 male Ss who appeared were used. No differentiation within this group was made on the kind of office for which they were candidates, since all the contested offices involved the responsibility of making decisions and influencing others.

To heighten the relevance of the cues of the ballot-counting situation to the story-writing procedure, the candidates were asked to judge their chances of winning the election, on the following scale: 0/5, 1/5, 2/5, 3/5, 4/5, 5/5 chances. The whole procedure was done anonymously.

The Ss in the *nonaroused condition* were 34 male students in an undergraduate course in psychology at the University of Michigan. The procedure was introduced as one about which some graduate students were interested in obtaining some normative data. After the procedure, the Ss were asked to sign their names, in order that further measures could be correlated with their story-writing performance. It seemed reasonable to regard this group as one for whom the concern about control of the means of influence was not specifically engaged.

One may argue that differences in respects other than aroused power motivation underlie such differences between these two groups as may be obtained in their stories. Perhaps

Ss in the aroused group have had peculiar learning experiences predisposing them to write power-related imagery for reasons having nothing to do with power motivation as defined. The validity of this argument is granted; there is no means of isolating this factor in the present design. Greater reliance is placed on the second step in the design—predicting power-motivated behavior from a measure derived from the first step—for establishing the tentative validity of the measure of power motivation. This problem will be more fully treated in discussion of the results.

Story-writing procedure. All Ss wrote stories for five pictures in a manner outlined in *The Achievement Motive* (8). The only procedural difference is that the pictures were reproduced in booklets rather than projected on a screen, and instructions modified accordingly. The five pictures are listed below in the order which they were presented:

1. Two men in a library; Picture E in the Achievement series (8).
2. Instructor in classroom; Picture 21 in Birney's study (5).
3. Group scene; Picture E in the Affiliation series (3).
4. Two men on campus; Picture 41 in Birney's study (5).
5. Political scene; from magazine, *Woman's Day*.

Step 2: Individual Differences in Power Motivation

Since the stories of the nonaroused Ss could be identified, other measures possibly related to a derived score of power motivation were obtained in this group: (a) Allport-Vernon Scale of Values (2), (b) Vocational Goals Inventory (7), (c) questionnaire on siblings, and (d) ratings by instructors of classroom behavior. The Vocational Goals Inventory required Ss to rank-order ten listed possible satisfactions on a job: being leader, interesting job, recognition from fellow men, being boss, surety of keeping job, self-expression, high pay, fame, helping others, working on one's own. The sibling questionnaire asked Ss to list their siblings and their ages, to permit reconstruction of birth order. The ratings by instructors pertained to the following classroom behavior: (a) Argumentative in the classroom—likes to argue and contradict; (b)

Eager to get points across in the classroom in order to convince either the instructor or other students of a point of view. They were made on a three-point scale of frequency: 1 = "very frequently," 2 = "at times," and 3 = "never or rarely." The following hypotheses were advanced concerning the relationship between individual differences in power-motivation scores derived from the projective instrument and the variables measured by the preceding measures:

Hypothesis 1. Power motivation is positively related to Political and Economic, and negatively to Social value orientation, as measured by the Allport-Vernon Scale of Values. A person highly interested in controlling the means of influence should value those ideas having to do with such control within the culture—namely, the political and economic values. Being interested in controlling a person in a sense negates his sympathy for this person's feelings, and therefore the negative relationship with the Social dimension was anticipated.

Hypothesis 2. Power motivation is positively related to the intensity of interests in the job satisfactions of being boss and being leader. These job satisfactions can be interpreted as positions in which the potentiality for controlling the means of influence is high.

Hypothesis 3. Power motivation is positively associated with the number of siblings in the S's family. Furthermore an association was hypothesized between birth order and strength of power motivation, though no prediction about the direction of the association is made. Power satisfactions are perhaps derived from many experiences in which being manipulated by others and manipulating others were in evidence. One would expect that the high power-motivated individuals would come from larger families where this type of experience is more likely to occur. Birth order would seem to be an important variable in further shaping these experiences.

Hypothesis 4. Power motivation is positively associated with instructors' ratings of argumentativeness and trying to convince others. These behaviors are presumably instrumental in satisfying the power motive in the classroom.

SCORING STORIES FOR POWER MOTIVATION

The method of scoring the stories for power motivation closely parallels the pattern established for hunger

(4), achievement (8), and affiliation (3) motives. The categories scored follow in a behavioral sequence analysis which is discussed with some detail in *The Achievement Motive* (8).

In scoring a story it is first identified as being related to power motivation or not. If a scorer decides that the story is related to power motivation (i.e., the story meets the Power Imagery criteria given below), it is then further scored for the presence or absence of the categories in the behavioral sequence. If the scorer decides that the story is not related to power motivation (Unrelated), no further scoring is done. The Power Imagery criteria are given in detail below. A complete scoring system with examples is available (13).

Power Imagery is scored when a story contains a reference to the thoughts, feelings, or actions of one of the characters which indicates that the character is concerned with the control of the means of influencing a person. Evidence of concern can come from one or more of three sources: (a) statements of affect; (b) statements of control activity; (c) statements of superior-subordinate role relations. These three criteria are elaborated below.

a. There is some statement of affect surrounding the maintenance or attainment of the control of the means of influencing a person. A character can be feeling good about winning an argument or feeling bad because he was unable to have his way. Also statements about wanting to win a point, or showing dominance, wanting to gain control (such as by a political or executive position), convince someone of something, or put a point across, can be interpreted as implicit statements of affective concern about the control of the means of influence. Affective concern can also be found in statements of wanting to avoid weakness. Examples of this are: being humiliated in a status position, being ashamed of an incapacity to assert one's self or become dominant, resenting the influence of another and wanting to overcome this.

Some very weak statements of concern over control of the means of influence are scored. Statements of desires to teach another person something, to interest another person in something, to inspire someone—although apparently weak in obvious power significance—should be scored. The only times when these statements would not be scored are in cases where the teaching, inspiring, or interesting is solicited by the person being influenced.

Statements of a person wanting another person to gain control of the means of influence—such as a person wanting someone else to win an election—should not be scored as Power Imagery. The only exception is a case in which there is a very clear identification of the person with the preferred candidate. This kind of imagery may appear in election stories where John wants to win the election, although John is not the candidate.

b. There is a definite statement of someone doing something about maintaining or attaining the control of the means of influence. Something that the character is actually doing in the story is the only kind of imagery that can qualify as Power Imagery under this criterion. The character has to dispute a position, argue something, demand or force something, try to put a point across, try to convince someone of something (and

theoretically any activity in order to obtain control of the means of influencing someone).

Three special considerations should be noted under Criterion *b*. The two special considerations noted under Criterion *a* are applicable to Criterion *b*. Statements involving trying to teach, interest, or inspire someone are scorable under Criterion *b*. Trying to win an election for someone else will not be scorable unless there is a close identification between the person campaigning and the candidate for office. Sometimes, moreover, it will be clear that an activity, although it meets the criterion listed in the preceding paragraph, is only for the purpose of arriving at the goal of another motive. In such a case, the story should not be scored for Power Imagery.

c. A story can be scored for Power Imagery if there is a statement of an interpersonal relationship which in its execution is culturally defined as one in which there is a superior person in control of the means of influencing a subordinate person. Examples of this are: boss-worker, judge-defendant. Mere mention of superior-subordinate relationship is not enough to score. There has to be

TABLE 1

NUMBER OF SUBJECTS IN AROUSED AND NONAROUSSED GROUPS ABOVE AND BELOW MEDIAN* FOR VARIOUS POWER-RELATED IMAGINATIVE CATEGORIES

Imaginative Category	Aroused Group (<i>N</i> = 34)		Non-aroused Group (<i>N</i> = 34)		<i>p</i> †
	Above‡	Below	Above	Below	
Power Imagery	27	7	15	19	.0023
Need	21	13	14	20	.047
Instrumental Activity					
Positive	22	12	15	19	.047
Negative	12	21	8	26	.090
Questionable	16	18	13	21	n.s.
Any sign	24	10	13	21	.0056
Goal Anticipation					
Positive	3	31	1	33	n.s.
Negative	3	31	3	31	n.s.
Any sign	6	28	3	31	n.s.
Block					
Personal	4	30	1	33	n.s.
Environmental	3	31	2	32	n.s.
Affective State					
Positive	4	30	1	33	n.s.
Negative	3	31	2	32	n.s.
Any sign	6	28	3	31	n.s.
Thema	28	6	17	17	.0044

* Median based on combined frequencies of both the aroused and the nonaroused groups.

† Except where indicated Ss were divided as to whether the category appeared once or not at all in the five stories, this procedure approximating the combined median most closely.

‡ All the probabilities were computed by Exact Test, one-tailed hypothesis, 1 *df*.

some mention of the activity involved in carrying out this relationship. Mere mention of a given person being influential would be enough to allow this story to be scored for Power Imagery.

Unrelated Imagery is scored when there is no evidence of power motivation concern in the stories as defined under Power Imagery.

When the story has been identified as one containing Power Imagery, it is scored for the presence or absence of the subcategories similar to those outlined for achievement motivation (8): Need; Instrumental Activity; Goal Anticipation; Block; Affective State; Thema.

In the final coding the coder did not know whether a given story came from one or the other of the two groups. To check on the reliability of this coding, a second coder scored a random set of twenty-four Ss from the total sample. The percentage of agreement for a given category between the two scorers ranged from 57 to 97. Furthermore, need for power scores (*n* Power scores) were tallied for each S by adding the number of categories appearing in his stories. The *n* Power scores obtained from the two coders were rank-ordered for the subsample. The rank-order correlation (*rho*) between the ranks obtained on these Ss by the two judges was +.87.

RESULTS

Step 1: Group Differences in Power Motivation

Imagery in the stories written by the aroused and nonaroused groups was compared by dichotomizing the frequency distributions of the combined groups as near to the median as possible. Table 1 gives for each category the number of Ss in each group above and below the median of the combined distribution. Exact tests were computed for each instance that potentially represented a significant deviation from a chance distribution. Power Imagery, Need, Instrumental Activity, and Thema appeared significantly more often in the aroused group's stories than in the nonaroused group's stories. These results, inci-

TABLE 2
NUMBER OF SUBJECTS IN AROUSED AND NONAROUSSED GROUPS ABOVE AND BELOW MEDIAN* OF *N* POWER SCORES

Group	<i>n</i> Power Score	
	Above 6	Below 6
Aroused group (<i>N</i> = 34)	22	12
Nonaroused group (<i>N</i> = 34)	9	25

Note.—*p* = .0013, one-tailed hypothesis, 1 *df*.

* The approximation to the median was computed by combining the *n* Power scores of both the aroused and the nonaroused groups.

TABLE 3
FREQUENCIES OF HIGH-LOW n POWER SCORERS IN VARIOUS CATEGORIES ON MEASURES
RELATED TO POWER MOTIVATION

Measure	Dimension	High n Power	Low n Power	Chi square	p	df
Allport-Vernon Scale of Values	Political					
	High*	8	5			
	Low	8	8	—	n.s.	—
	Economic					
	High	6	5	—	n.s.	—
	Low	10	8			
Vocational Goals Inventory	Social					
	High	7	9	—	.12	1
	Low	9	4			
	"A job where you could be leader"					
	High	9	4	—	.096†	1
	Low	7	10			
	"A job where you could be boss"					
	High	8	8	—	n.s.	—
	Low	8	6			
	"A job where you could be looked upon very highly by fellow men"					
Sibling Questionnaire	High	10	3	—	.049	1
	Low	6	11	(2 tailed hypothesis)‡		
	No. of Siblings					
	0	0	1			
	1	7	4	—	n.s.	—
	2	4	3			
	3 or more	5	3			
Instructor's Ratings	Has older sibling					
	Yes	9	5	—	n.s.	—
	No	7	6			
	Argumentativeness					
	Frequency					
	High	7	1			
	Medium	7	7	5.62§	<.05	2
	Low	4	7			
	Frequency of Trying to Convince Others					
	High	9	1			
	Medium	7	6	11.02	<.001	2
	Low	2	8			

* High and Low division on Allport-Vernon Scale of Values based on deviation from normative data (2).

† The probabilities of the 2 × 2 tables were computed by the Exact Test.

‡ All other probabilities were estimated under a one-tailed hypothesis.

§ Chi squares where reported were estimated by the Mood Maximum Likelihood Ratio (9).

dentally, are comparable to those obtained in similar comparisons in respect to the achievement and affiliation motives, except that the categories Goal Anticipation, Block, and Affect failed to differentiate the groups.

An over-all check on the differences in imagery can be made by comparing the n Power scores for the Ss in the two groups. Table 2 shows the number of Ss in each group above and below the combined groups' median n Power score. The probability of obtaining this distribution by chance is .0013, again as estimated by an Exact Test.

Step 2: Individual Differences in Power Motivation

Scores of the nonaroused group for n Power were dichotomized as near as possible to the group median. The resulting division into High and Low n Power scores is retained through all comparisons to be made in this section. The results of the comparisons made between the n Power scores and the variables for which relationships were predicted appear in Table 3. In most cases the other variables are dichotomized near the median into high and low categories.

This table supports the following observations which confirm predicted hypotheses. High *n* Power scorers, as compared to low *n* Power scorers, tend to have lower scores on the Social Value Dimension of the Allport-Vernon Scale of Values; show a significantly stronger interest in the job satisfaction of being leader; were rated significantly higher by their instructors on the frequency of argumentation and the frequency of trying to convince others of their points of view in the classroom.

Other hypotheses were not confirmed. High *n* Power scorers as compared to low *n* Power scorers did not differ significantly in their scores on the value orientations Economic and Political, on their ratings of the job satisfaction of being boss, and on the number or order of siblings in their families.

Table 3 also includes an unpredicted result, that high *n* Power scorers rated the job satisfaction of obtaining recognition from their fellows significantly higher than low *n* Power scorers. No other comparison between the motive measure and the Vocational Goals Inventory yielded any positive results.

DISCUSSION

The results obtained comparing the imagery produced by the aroused and nonaroused groups are similar to parallel results found with the measures of achievement and affiliation motivations. In the present design however there is an ambiguity in the nature of the arousal condition which was not present in the earlier studies. This ambiguity—the fact that the power arousal group was different from the nonaroused group not only in presumed level of arousal of some motivational pattern but also in whatever dispositions are salient in discriminating candidates and noncandidates for college offices—forces one to make only tentative conclusions about these results. It may be that the group differences uncovered in the analysis of the stories are due to motivations which are distinct from the power motive as defined.

Two other reasonable interpretations of the obtained differences between the two groups attribute them to differences in the level of achievement or recognition motivation. Fortunately it was possible to check the former possibility. The two sets of stories were scored for achievement motivation as outlined in

The Achievement Motive (8). No significant difference in mean *n* Achievement scores of the groups was found.

The other interpretation, that the group differences are accountable in terms of recognition rather than power motivation, is a distinct possibility. The only basis for thinking otherwise is that the scoring system was built around a conceptualization of power motivation as distinct from strivings for recognition. Recognition themes in stories were not scored unless they fulfilled the criteria for power motivation as well. That is, if a story brought up problems of recognition as a way of obtaining control of the means of influence, then and only then was it scored. There is, however, no experimental way in the present study to rule out the interpretation of the power-motive measure in terms of recognition. In further use of the measure of power motivation caution to recognize this possible contamination is recommended.

There seemed to be no difference between the aroused and nonaroused groups on certain of the affected categories (Goal Anticipation, Affect, and Obstacle), which proved valuable in discriminating groups with the other measures. An explanation for this discrepancy should be offered. Perhaps this particular motive does not permit the arousal of associated affective thoughts. There might be good cause to inhibit the expression of affect surrounding the power area, since there are many recognized taboos for expressing power in this culture—especially the expression of overt satisfactions derived from power behavior. Support for this explanation comes from the observation that these categories appeared very infrequently in the stories of both groups. Another reason for the lack of differential appearance of these categories might be that the particular pictures used in this study did not provide cues to the kind of thinking that would reflect the affective component. This possibility can be tested in further research.

The fact that the measure not only reflects the group differences in assumed motivation level, but also individual differences in variables related to aspects of power concern as defined, suggests that the instrument can be used with success in researches in which such a measure of motivation would be needed. The measure of power motivation, if some of its

ambiguities are clarified, can be potentially used for exploring dimensions of control behavior exhibited in interpersonal relationships. Researchers using concepts of power different from the one presented here, should be cautious in the use of this measure.

Some of the hypothesized relationships, particularly with the Allport-Vernon scales, were not substantiated in this study. These value measures can perhaps reflect many different motivational orientations, and it may thus be naive to expect a positive relationship between a fairly specific motivational pattern and values that can service many different needs.

Lack of support for the prediction that Ss high in *n* Power would rate the job satisfaction of being boss higher than Ss low in *n* Power may indicate that re-evaluation of the meaning of being boss for them is in order. "Boss" can be thought of as a negative word in this culture, and, indeed, there was little variability in the Ss' ranking of job satisfaction in this respect. Nearly everyone ranked this alternative low.

The simple prediction that high *n* Power Ss would come from large families was not substantiated. This finding would seem to indicate that any relationship between the power motive and the family structure depends on more complex considerations than simple frequency of contact with other family members. When the distribution of *n* Power score is divided into thirds (high, medium, and low), a trend was uncovered. Eight out of twelve high *n* Power scorers had at least one older sibling whereas this was true for only one out of six medium *n* Power scorers and five out of nine low *n* Power scorers. (Chi square = 4.34, $p < .15$, 2 *df*, two-tailed hypothesis.) Perhaps some fruitful distinctions can be made in the area of family structure which may be useful in understanding the development of the power motive.

The finding that high *n* Power Ss rated the job satisfaction of gaining recognition from their fellow men higher than do low *n* Power Ss points to the possibility that the satisfactions of prestige may be intimately tied with the satisfactions of power. This point of view has been elaborated by Murphy (9). There is also the possibility that the measure includes, at least partially, some aspects of recognition motivation. Indeed, all of the significant findings can be theoretically accounted for by

considering the motive measured to be the recognition motive.

The relationship between achievement and power motivation. The product-moment correlation between the *n* Achievement and *n* power scores was $+ .27$ ($p < .05$). An attempt was made to use the *n* Achievement scores, however, to predict the same variables with which *n* Power scores had been related. No significant findings with these variables were obtained. In short, there seems to be a slight positive association between the two motives, but what is measured in scoring for the power motive is that process which seems to be directly related to power, rather than achievement strivings.

SUMMARY

Two groups of Ss wrote stories in response to five pictures. One was tentatively considered to be aroused by cues relevant to power motivation; the other was not aroused. A scoring system paralleling the coding methods which had been fruitful in measuring the achievement and affiliation motives was applied to the stories, and succeeded in differentiating the two groups. Four of the seven categories coded appeared significantly more frequently in the stories of the aroused group.

Individual differences in power motivation were estimated for Ss in the nonaroused group by the *n* Power scores (based on the coding categories). High *n* Power scorers differed from low *n* Power scorers in this group on a number of other factors relevant to power motivation. Some hypotheses relating strength of power motivation to other variables were not confirmed.

Although there was a significant positive correlation between *n* Power scores and *n* Achievement scores, the *n* Achievement score distribution does not predict the same relationships with power-related variables.

In conclusion, the results indicate that the projective measure of power motivation developed here not only can successfully isolate the presumed differential level of motivation in groups, but also can successfully predict attitudes and overt behavior which are presumably related to the processes involved in power motivation. It is cautioned, however, that the measure may be conceived also in terms of recognition motivation.

REFERENCES

1. ADLER, A. *Understanding human nature*. New York: Greenberg, 1927.
2. ALLPORT, G. W., & VERNON, P. E. *Manual for the study of values* (Rev. Ed.) Boston: Houghton Mifflin, 1931.
3. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405-410.
4. ATKINSON, J. W., & MCCLELLAND, D. C. The projective expression of need. II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643-658.
5. BIRNEY, R. Studies on the role of picture cues in projective measurement of achievement motivation. Unpublished doctor's dissertation, Univer. of Michigan, 1954.
6. CARTWRIGHT, D. Toward a social psychology of groups. Presidential address delivered before SPSSI, Cleveland, Ohio, 1953.
7. CENTERS, R. Motivational aspects of occupational stratification. *J. soc. Psychol.*, 1948, **28**, 187-217.
8. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R., & LOWELL, E. L. *The achievement motive*. New York, Appleton-Century-Crofts, 1953.
9. MOOD, A. M. *Introduction to the theory of statistics*. New York, McGraw-Hill, 1950.
10. MURPHY, G. Social motivation. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 601-633.
11. SHIPLEY, T. E., JR., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349-356.
12. SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William Alanson White Psychiat. Found., 1947.
13. VEROFF, J. Development and validation of a projective measure of power motivation. Unpublished doctor's dissertation, Univer. of Michigan, 1955.

Received February 27, 1956.

GSR REACTIVITY AS A FUNCTION OF ANXIETY, INSTRUCTIONS, AND SEX

JAMES L. BERRY AND BARCLAY MARTIN

University of Wisconsin

SEVERAL studies indicate that GSR conditioning is positively related to the level of anxiety of the subject. Bitterman and Holtzman (4) found such a relationship when an over-all index of anxiety was assessed on the basis of several behavioral ratings and test scores. Welch and Kubis (10) found that psychiatric patients diagnosed as anxious showed more rapid GSR conditioning than a group of normals. Beam (3) found GSR reactivity to increase immediately before anxiety-arousing situations, such as oral examinations, as compared to a measure taken at another time. However, results have been rather uniformly negative in terms of a relationship between anxiety and GSR conditioning when the Taylor Anxiety Scale has been the sole measure of anxiety level. Neither Bitterman and Holtzman nor Beam found any relation between GSR conditioning and the Taylor Anxiety Scale.

The present study was designed to test the hypotheses that the Sarason Test Anxiety Scale and the kind of instructions given to the Ss are related to GSR conditioning. The Test Anxiety Scale developed by Sarason (6) was tried in this study because of the previous failures of the Taylor Scale to be predictive in this area, and because it was thought on an a priori basis that the nature of the Sarason Test was such as to give a more valid measure of anxiety proneness among college students. The Test Anxiety Scale requires S to indicate on a series of five-point scales how he feels right before and during examinations in college courses, group intelligence tests, and individual intelligence tests.

Three different kinds of instructions were given with the intent of varying experimentally the level of apprehension that the Ss would have about the conditioning situation. Taylor (9) varied instructions in a conditioned eye-blink study by telling one group that the air puff would increase in intensity as the trials progressed and telling another group that it would decrease. She found no differences between the groups. No previous attempts to

modify GSR conditioning by instructions were discovered.

METHOD

From a class of 330 introductory psychology students, extreme groups of 60 high-anxiety (30 male and 30 female) and 60 low-anxiety (30 male and 30 female) Ss were selected. The Ss from each of these extreme groups of the same sex were then placed randomly in one of the three instructions groups. This procedure provided an anxiety-by-instructions-by-sex factorial design with 10 Ss in each of the twelve groups.

A modified form of the Sarason Test Anxiety Scale,¹ with items referring to individual intelligence tests omitted, had been administered to the class as a group. Answers were recorded on regular five-point IBM answers sheets. The Ss were then seen individually for the second session.

The Apparatus used to measure GSRs was an AC circuit designed by Grant (5). Dry silver finger electrodes were held on the palmar surfaces of two adjacent fingers on the right hand by rubber straps. Shock was supplied by a vacuum tube circuit which provided a constant current of .30 ma. to the left arm of the S. An audio-oscillator provided a 4,000 cycle tone. Cams attached to a one-cycle-per-minute motor were constructed so that a tone of 5 seconds' duration followed immediately by a shock of 0.5 second's duration could be presented every 30 seconds. The shock could be eliminated when desired by a switch.

The instructions given to the three different groups were as follows:

Group 1, Apprehension Arousal. Throughout the course of this experiment you will receive several electric shocks, such as the following (2 sample shocks). At one point in the experiment you will receive a shock considerably stronger, approximately twice as strong as the ones you just experienced. We realize that the shock is already strong and, because increasing it too much might have bad effects, you will receive only one of the stronger ones. This one may occur at any time during the course of the experiment. Do you have any questions before we begin?

Group 2, Neutral. Throughout the course of this experiment you will receive several electric shocks, such as the following (2 sample shocks). Do you have any questions before we begin?

Group 3, Reassurance. Throughout the course of this experiment you will receive several electric shocks, such as the following (2 sample shocks). These may seem rather strong to you now but you will find that you will adapt to them rather quickly and, after a few more, it will begin to feel like a pricking or

¹ We are grateful to Dr. Sarason for supplying us with a copy of the Test Anxiety Scale.

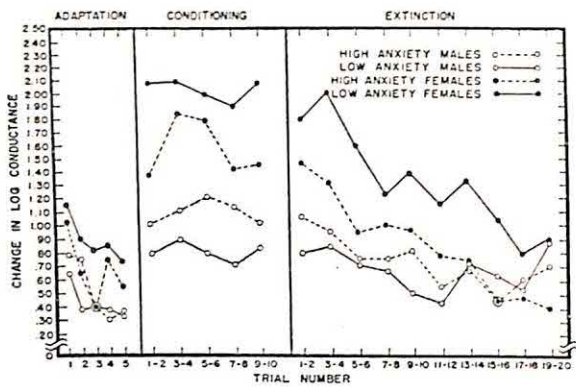


FIG. 1. ADAPTATION, CONDITIONING, AND EXTINCTION OF MALE AND FEMALE, HIGH- AND LOW-ANXIETY GROUPS

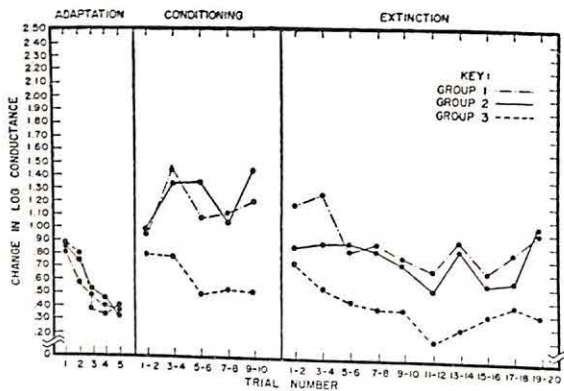


FIG. 2. ADAPTATION, CONDITIONING, AND EXTINCTION OF MALE INSTRUCTION GROUPS 1, 2, AND 3

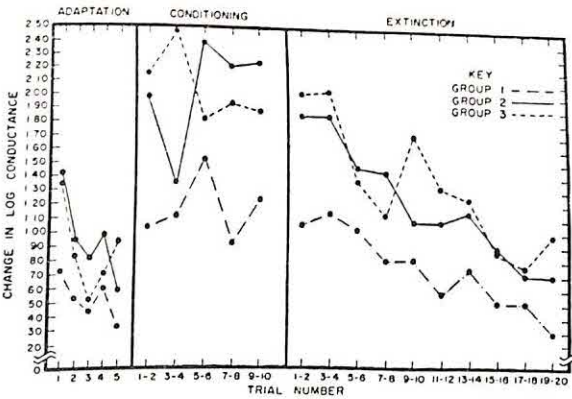


FIG. 3. ADAPTATION, CONDITIONING, AND EXTINCTION OF FEMALE INSTRUCTION GROUPS 1, 2, AND 3

tickling sensation. Most subjects haven't been bothered by them. Do you have any questions before we begin?

Each *S* was then given identical treatment in the conditioning situation. Four minutes of relaxation were followed by five trials with the tone alone, after which 10 conditioning trials and 20 subsequent extinction trials were administered. A partial reinforcement schedule was used during conditioning in which no shock was given on trials, 3, 5, and 9. Skin resistance

TABLE 1
SUMMARY OF THE ANALYSIS OF VARIANCE OF
CONDITIONING TRIALS

Source	df	Mean Square	F	p
Total	119			
Within Groups	108	7.89		
Sex	1	103.9	13.2	.01
Anxiety	1	.29	.04	
Instructions	2	14.8	.93	
Sex × Anxiety	1	23.81	3.01	.10
Sex × Instructions	2	26.85	3.40	.05
Anxiety × Instructions	2	4.44	.56	
Sex × Anxiety × Instructions	2	1.92	.37	

readings were taken immediately before the tone and of the lowest point reached during the tone for all trials.

RESULTS

The curves for adaptation, conditioning, and extinction are shown in Figs. 1-3. Resistance was transformed to log conductance - 1 to minimize the correlation between GSR and base level that is found when resistance is used as the measure, and trials were combined in pairs for conditioning and extinction curves. A summary of the analysis of variance for repeated measures for the average response during the conditioning trials is shown in Table 1. Bartlett's test for homogeneity of variance was nonsignificant. The *F* values for sex and the interaction of sex and instructions were significant at the .01 and .05 levels respectively. The interaction of sex and anxiety did not reach statistical significance.

In order to investigate the possibility of consistent trends as a function of trials, an Alexander trend analysis (1) was performed. The over-all group slope was not significantly different from zero, nor were the subgroups significantly different from each other in slope. It is apparent from Fig. 1 that most of the increased reactivity during the conditioning trials as compared with the adaptation trials occurred during Trials 1 and 2.

The extinction data were analyzed by computing the average response of each *S* over the 20 extinction trials, and then performing an analysis of variance upon these scores. Table 2 summarizes this analysis, and indicates that only the between-sex variance was significant.

² In order to make the figures easier to read, the transformation was actually as follows: 100 (log conductance - 1).

TABLE 2
SUMMARY OF THE ANALYSIS OF VARIANCE
OF EXTINCTION TRIALS

Source	df	Mean Square	F	p
Total	119			
Within groups	108	99.1		
Sex	1	502.7	5.07	.05
Anxiety	1	150.5	1.52	
Instructions	2	31.4	.32	
Sex $\times$ Anxiety	1	254.6	2.57	
Sex $\times$ Instructions	2	253.7	2.56	
Anxiety $\times$ Instructions	2	57.4	.58	
Sex $\times$ Anxiety $\times$ Instructions	2	16.6	.16	

The sex difference was found to hold during adaptation also. The average of Trials 4 and 5 of tone alone was found to be significantly higher for the female group ($F = 14.1$, $p < .001$).

The results are more clearly visualized in the curves. A comparison of the male and female curves show a marked tendency for females to have greater GSRs during all phases of the experiment. The interaction of sex by instructions is apparent when Fig. 2 and 3 are compared. For the males, the Reassurance instruction group is considerably below both of the other instruction groups. For the females, the Apprehension Arousal group is the lowest of the three. Thus, there is a tendency for the instructions to have opposite effects on males and females. This tendency was further analyzed within each sex group, using the average response during conditioning as the measure. The between instructions variance was significant for the female groups ($F = 3.59$, $p < .05$). The between-instructions variance for the male groups, however, was not significant ($F = 2.44$, $p < .10$). It is apparent from the curves that the Apprehension Arousal and Neutral male groups are very similar in their conditionability and that the Reassurance group falls below these groups. Since it was expected that the average response during conditioning would be less for the Reassurance group, it was considered appropriate to compute the probability of this occurrence. This was done by the use of Whitney's extension of the U statistic as described by Mosteller and Bush (7). The probability of the Reassurance group's being lower than both the other groups by chance alone was less than .01.

Although the term "conditioning" has been used throughout this study, it is probable that much of the variance may be the result of individual differences in general GSR reactivity, therefore reflecting pseudo-conditioning to some extent. That this is probably the case is suggested by other GSR data recently obtained in the Wisconsin laboratory on an N of 80 Ss in which the average GSR reactivity during trials with tone alone correlated .74 with average response during conditioning and .64 with average response during extinction. The possibility that pseudo-conditioning is involved of course does not make possible relationships between GSR reactivity and anxiety or instructions any less meaningful.

The marked sex differences in GSR conditioning found in this study are consistent with a similar finding with eye-blink conditioning reported by Spence and Farber (8). Females gave more conditioned responses ($p < .10$) than males. Similarly, Aronfreed, Messick, and Diggory (2) found greater GSR reactivity in females than males. Whether females are constitutionally more reactive in GSRs or more apprehensive in situations of this sort obviously cannot be determined from these results.

The findings seem to support the hypothesis that instructions affect the extent of GSR conditioning. Reassuring instructions to the male Ss resulted in lessened conditionability. The significant interaction of sex and instructions is puzzling, and suggests that whereas the male groups are affected in the predicted and common-sense direction by instructions, the female groups are affected oppositely, Apprehension Arousal instructions producing a decrease in conditionability. A similarly puzzling tendency, although not significant, was for the female high-anxiety group to show less conditioning than the low-anxiety group. It is perhaps best to withhold speculation about this contrariness of females until such an unexpected finding can be cross validated.

The Sarason Test Anxiety Scale did not prove to be significantly related to GSR conditioning. Along with the previous failures of the Taylor Anxiety Scale to be related to GSR conditioning, this finding raises the question of whether any set of paper-and-pencil items is related to GSR conditioning. This question is being presently pursued by empirical item analysis.

SUMMARY

GSR conditioning and extinction measures were obtained on 120 college Ss in a factorial experimental design in which twelve treatment groups were divided into two sex groups, three instruction groups, and high- and low-anxiety groups. The three instruction groups were given three different kinds of instructions designed to be (a) apprehension arousing, (b) neutral, and (c) reassuring. The high- and low-anxiety groups were selected as extreme groups on the basis of the Sarason Test Anxiety Scale.

Results indicated that the instructions had opposite effects on the conditionability of males and females. The male Ss tended to show less conditioning when given reassuring instructions than when given either neutral or apprehension arousing instructions. The female Ss tended to show the least conditioning with the apprehension-arousing instructions and most conditioning with the reassuring instructions.

There was no relationship between GSR conditioning and the Sarason Test Anxiety Scale. Females showed greater GSR reactivity than males in all phases of the experiment; adaptation, conditioning, and extinction.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533-557.
2. ARONFREED, J. M., MESSICK, S. A., & DIGGORY, J. C. Re-examining emotionality and perceptual defense. *J. Pers.*, 1953, **21**, 517-528.
3. BEAM, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543-551.
4. BITTERMAN, M., & HOLTZMAN, W. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615-623.
5. GRANT, D. A. A convenient alternating current circuit for measuring GSR's. *Amer. J. Psychol.*, 1946, **59**, 149-151.
6. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166-173.
7. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 317-319.
8. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116-119.
9. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
10. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83-91.

Received February 2, 1956.

THE RELATIONSHIPS AMONG IMAGINATIVE, DIRECT VERBAL, AND PHYSIOLOGICAL MEASURES OF ANXIETY IN AN ACHIEVEMENT SITUATION¹

ALFRED C. RAPHELSON²

Bureau of Educational Research, University of Illinois

SEVERAL independent attempts have been made to relate different measures of anxiety (or fear) to performance in competitive achievement situations. In one series of studies (3, 11, 12, 15), anxiety level has been assessed by the Manifest Anxiety Scale (MAS), composed of refined MMPI items which have been classified by clinical psychologists as being indicators of general manifest anxiety. The experimental results indicate that subjects who score high in manifest anxiety condition faster in a classical type of experiment than low-scoring subjects, but show poorer performance in more complicated learning situations.

Mandler and Sarason (7) have developed a Test Anxiety Questionnaire (TAQ) which indicates the extent to which a subject experiences anxiety before, during, and after three typical kinds of testing situations, thus having more specific situational reference than the MAS. Using the TAQ, Sarason, Mandler and Craig-hill (8) obtained results indicating that as the expectation of failure is increased, the high-anxiety subjects show a decrement in performance, whereas the low-anxiety subjects show an increment.

McClelland *et al.* (6) placed subjects in situations differing in aroused motivation to achieve and asked them afterward to write stories to TAT-like pictures. From these stories a scoring system was developed to capture the thought content reflecting the degree to which a subject was motivated to excel in an achievement situation. This need-Achievement (n Ach) score related to behavior in a way that suggests that qualitative differences in thought content of persons who score high and low in n Ach represent real differences in expectancies concerning competitive situations

(1). McClelland (5) speaks of the high scorers as showing a positive hope of success. He infers a defensive fear of failure from the relative absence of achievement imagery and relative preponderance of negative categories in the stories of the low scorers.

The present study attempts to investigate the interrelationships among these three measures and to test their relative effectiveness in accounting for reactions at another level of observation, i.e., physiological activity. All three measures give rise to distributions of scores described in terms of the amount of fear or anxiety that characterizes an individual generally (MAS) or specifically in competitive achievement situations (TAQ, n Ach). Traditionally, increases in skin conductance have been found to be correlated with emotional states. It is expected, therefore, that groups characterized as anxious on the individual difference measures (high MAS, high TAQ, low n Ach), if validly classified, will show conductance changes that are more consistent with the patterns expected in an emotional state than the groups classified as nonanxious (low MAS, low TAQ, high n Ach).

METHOD

Apparatus

The data were obtained in a room, 9 × 9 × 10 feet, in which was located the complex structure housing the essential apparatus. These essentials consisted of (a) a ground-glass screen 24 inches in diameter on which a 3/4-inch spot of light moved, (b) a device which provided for a very irregular pattern of movement of this spot, (c) a recording polygraph, (d) a galvanometer, (e) a pneumograph, and (f) a red signal light.

The movement of the spot was accomplished by the projection of a 3/4-inch spot on the screen by a mirror arrangement which was driven by two motor-driven cams in a highly irregular manner. A hand lever was connected in opposition to this device. A larger circle (3 inches in diameter) was also projected on the screen.

The recording galvanometer was used in conjunction with an electric network which automatically kept constant the current passing through the skin of S at 40 microamperes and measured the apparent resistance in terms of the potential difference across the electrodes. A continuous record of the changes in resistance of the skin was maintained by connecting the galvanom-

¹ This article is a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy to the faculty of the University of Michigan. The writer wishes to express his deep gratitude to Drs. John W. Atkinson and Carl R. Brown for their patient direction and cooperation.

² Now at University of Michigan, Flint College.

eter to a stylus on the polygraph. One electrode was attached to the palm of the left hand, the other to the underside of the wrist on the same hand. Each electrode was made of a fritted glass filter of fine porosity, filled with physiological salt solution in which a loose coil of anodized silver wire was emersed.

The pneumograph was strapped to the abdominal region of *S* and connected to the polygraph.

A red signal light, located above the screen, indicated when *S* had allowed an error to occur. The errors were also indicated on the polygraph.

Subjects

The *Ss* were 25 volunteer male students enrolled in various schools in the University of Michigan. Their class status ranged from freshman to graduate, and their age from 18 to 25.

General Plan

The general plan was to establish a testlike situation in which *S* was asked to keep the moving spot of light within a small area by manipulating the hand lever. In a brief introduction, the apparatus was described as being like devices used by the Air Force to test intellectual and motor abilities. The following test instructions were given:

I am going to give you a series of reasoning problems that I want you to solve as best you can. I am not going to ask you for a solution; I do not care whether or not you can verbalize it. But I want you to show me you know the answer by your responses. Each problem consists of a sequence of movements of this spot of light over the screen. I will start the machine which controls the movements of the spot and your task is to discover what the sequence of movements is that the light follows. Now the leverstick on your right works in opposition to the machine. If the machine moves the spot up you can counter the movement by moving the lever so that the spot moves in the opposite direction. Therefore, if you are able to solve the problem correctly, you will be able to anticipate each subsequent movement of the light, and by manipulating the lever correctly, you can keep the spot from moving very far. In the center of the screen you can see a ring outlined. Your task is to keep the spot of light from wandering out of the limits of this ring. It is an error if you allow the spot to leave that area. Your task is to make a good score by making as few errors as you can. The errors will be recorded on my tape here and after the session we will compare your score to that of others who have been through this test. You are to try your best to make as few errors as possible.

Now in this problem you will have several other cues to aid you. First, the ring will always be there to indicate where the safe-area ends and the error-area begins. Second, every time you make an error this red light above the screen will flash on. It not only will indicate that you are in error but will also aid you in keeping a running count as to how well you are doing. Before you begin the test, I am going to begin the sequence, and I want you just to observe it for awhile so you can see what the problem is like and try to figure out the pattern of movements. I will tell you when I will begin the sequence and

when you are to begin avoiding making errors. Do you have any questions?

Continuous records were taken of skin conductance and respiration during the entire experimental session.³ The session consisted of the following four periods: *Null 1 Period*—a five-minute period in which *S* actively manipulated the lever but was not yet given the test instructions. *Anticipation Period*—a two-minute period which immediately followed the test instruction. The *S* watched the movements of the spot but did not yet begin to avoid making errors. *Solution Period*—a ten-minute period in which *S* actively attempted to make as few errors as he could by controlling the movements of the spot. *Null 2 Period*—a five-minute period after the Solution Period which was similar to the Null 1 Period. No problem was presented while the *S* simply manipulated the lever. The first and last periods were conceived of as relatively free of anxiety-arousing cues. It was not expected that the subgroups differentiated by the various dispositional measures would be different in their physiological reactions in these periods.

Since the spot actually moved in an irregular manner and no solution was possible, errors are not a meaningful variable in this experiment. An attempt was made to keep the frequency of errors as constant as possible for each *S* by manipulating the amplitude of the spot's movement in relation to *S*'s actual performance.

Measures of Individual Difference

Following the motor task, *Ss* wrote the *n*-Ach stories while still connected to the recording equipment. They were disengaged before taking the other two tests. To be consistent with the literature on these measures, each distribution of scores was broken into high, middle, and low thirds for comparison on the physiological indices. The *n*-Ach measure consisted of four TAT-like pictures, presented to elicit imaginative responses from which a score was determined according to the coding procedure described by McClelland, *et al.* (6). The experimenter and another trained coder scored all the stories. The reliability coefficient between the two sets of scores was .82. Differences were resolved by reference to a third expert.⁴ The scores for the high-*n*-Ach third ranged from 8-14; for the low third, -3-2.

The Test Anxiety Questionnaire has been fully described by Mandler and Sarason (7). A high score indicates a tendency to reflect feelings of anxiety in a testing situation. The scores for the high-TAQ third ranged from 23-31; for the low-TAQ third, 1-13.

The Manifest Anxiety Scale has been described by Taylor (14). The scores for the high-MAS third ranged from 15-22; for the low third, 1-7.

Skin Conductance Measure

The skin resistance records were analyzed in the following way: Samples from the records were taken

³ No significant relationships were found between any of the other variables in these experiments and any of the characteristics of respiration activity (amplitude, rate, volume). Hence no respiratory data will be reported.

⁴ The writer is indebted to Mr. Berne Jacobs and Dr. Joseph Veroff for their cooperation in this scoring.

at the beginning of each period and at the end of each subsequent minute. These readings were converted into units of conductance (micromho) and an average based on these samples was computed to represent the entire period. Analyses were done on the average level in each period in terms of variation in conductance within each period expressed as a percentage change over the Null 1 Period average. The percentages were calculated as follows:

$$\frac{\text{Mean Conductance in period}}{\text{Mean Conductance in Null 1}} \times 100.$$

These percentages were converted into angles by the arc sin $\sqrt{\text{percentage transformation}}$ to insure the independence of means and variances (10).

Statistical Analysis

The error estimates used to evaluate the mean differences were derived from analyses of variance designated Type III by Lindquist (4). Since there were twenty-five Ss in this study, it was impossible to have equal numbers in each of the three subgroups for each measure. The numbers in each subgroup, therefore, were equalized by randomly eliminating one case from the subgroup that contained the unequal number.

RESULTS

Interrelationships Among Measures of Individual Differences

The product-moment correlations among the three measures of individual differences are given in Table 1. The two measures of reactions specific to achievement (n Ach, TAQ) were significantly correlated, as were the two measures concerned with negative reactions towards general situations (MAS, TAQ). This pattern of relationships is consistent with the specifications set for the measures by the theories to which they are related. The TAQ was developed to measure anxiety symptoms specific to an achievement situation. The MAS was developed to measure general feelings of anxiety which may or may not occur in an achievement situation. A significant positive correlation is expected between them. The n-Ach measure is thought to reflect the degree to which S has expectancies of goal attainment in an achievement situation; such expectancies are incompatible with any anxiety reactions associated with the setting. A high negative correlation is expected with TAQ and a low negative correlation with MAS.

Absolute Conductance Level in Null 1 Period

The initial Null 1 levels of conductance in relation to the several measures of individual

TABLE 1
PRODUCT-MOMENT CORRELATIONS AMONG
THREE MEASURES OF INDIVIDUAL
DIFFERENCES

Measure	Test Anxiety	N	Manifest Anxiety	N
n Achievement	-.43*	24†	-.25	25
Test Anxiety			.53	24

* r 's of .39 and .50 are significant at .05 and .01 levels of confidence respectively.

† One subject failed to fill out the Test Anxiety Questionnaire correctly and was eliminated from all Test-Anxiety analyses.

TABLE 2
MEAN ABSOLUTE CONDUCTANCE IN NULL 1
PERIOD

Measure	High (n = 8)	Middle (n = 8)	Low (n = 8)	σ diff.
n Achievement	23.9*	20.7	17.5	6.42
Test Anxiety Questionnaire	17.3	22.2	23.5	6.40
Manifest Anxiety Scale	17.7	22.4	22.0	6.35

* Units in micromhos.

TABLE 3
MEAN VARIATION IN CONDUCTANCE EXPRESSED
AS PERCENTAGE OF NULL 1 LEVEL
(N = 8 for all subgroups)

Measure	Subgroup	Period		
		Anticipation	Solution	Null 2
n Achievement	High	102.5*	95.4	68.9
	Middle	101.1	106.5	85.0
	Low	100.1	110.0	79.4
Test Anxiety Questionnaire	High	104.2	109.2	79.6
	Middle	101.0	103.1	86.8
	Low	103.8	89.5	66.8
Manifest Anxiety Scale	High	106.0	109.6	82.8
	Middle	94.0	98.8	71.2
	Low	103.2	101.8	78.6

* The σ diff. for each measure is as follows: n Ach: between = 16.4, within = 7.4; TAQ: between = 16.3, within = 6.4; MAS: between = 16.5, within = 7.6.

differences are presented in Table 2. In this period, the anxiety-arousing instructions had not yet been introduced. However, on each of the three measures, the subgroup which is presumed to be composed of relatively nonanxious Ss showed the highest conductance level. Although even the largest difference, that obtained between the subgroups scoring high and low on n Ach, was not significant (σ diff. = 6.42, $t = .997$), the consistency of the trends in each of the comparisons may be worth noting. They

TABLE 4

MEAN VARIATION IN CONDUCTANCE EXPRESSED
AS A PERCENTAGE CHANGE OF NULL 1
LEVEL FOR SUBJECTS DESIGNATED
ANXIOUS OR NONANXIOUS*

Subgroup	N	Period		
		Anticipation	Solution	Null 2
Anxious	7	104.4	112.1	83.7
Nonanxious	9	101.7	89.0	67.6
Difference		2.7	23.1	16.1
σ diff.		12.42	12.22	9.05
t		.22	1.89	1.78
p (one-tailed)		—	.04	.05

* On the basis of scores on both the Test Anxiety and the achievement measures.

are consistent with the notion that a high conductance level may be indicative of psychological states other than anxiety.

Skin Conductance Variation in Experimental Periods

The mean variation in conductance, expressed as a percentage change over Null 1 level in three experimental periods, is presented in Table 3 for each of the subgroups differentiated by the three measures. The subgroups differentiated by measures of n Ach and TAQ showed significantly different conductance patterns. Though none of the between-subgroup differences were significant at any particular period, there were meaningful differences between subgroups in the direction of the trend of skin conductance through the different experimental periods. The subgroups considered relatively "anxious" (low n Ach, high TAQ) showed an increase in conductance from the Anticipation to the Solution Period. The subgroups considered relatively nonanxious (high n Ach, low TAQ) showed a decrease. The differences between these particular differences with both measures were significant in the predicted direction (n Ach: diff. = 17.0, σ diff. = 10.5, $t = 1.62$, $p = .05$; TAQ: diff. = 19.3, σ diff. = 9.1, $t = 2.12$, $p = .025$).

The subgroup designated as anxious by the more general MAS did not show any pattern of change in conductance that was significantly different from the subgroup designated as nonanxious.

A final analysis was made, comparing only those S s scoring both high in TAQ and low in n Ach (both treated in the past as symptomatic

of a disposition to be anxious in achievement situations) with S s who scored both low in TAQ and high in n Ach (both treated in the past as symptomatic of the relative lack of a disposition to be anxious in an achievement situation). The significant negative correlation between the two measures implied that by examining S 's score on both measures, it would be possible to obtain a more reliable estimate of the degree to which he might be anxious or nonanxious in a competitive situation. On each measure, the distribution was divided as near as possible to the median. Only S s who fell in the corresponding high and low anxiety halves of both distributions were selected for comparison. Mean variations in conductance for these subgroups are presented in Table 4. It can be seen that by this method of classifying S s, the trends observed between subgroups in the separate analyses were further emphasized and were significant in the Solution and Null 2 Periods. The directional change between the Anticipation and Solution Periods observed, significantly different in the separate analyses, was also significant here (diff. = 20.4, σ diff. = 10.1, $t = 2.01$, $p = .03$).

The trends observed in the earlier analyses concerning the initial level of conductance in Null 1 were further emphasized in this method of classification. The subgroup designated nonanxious showed a significantly higher initial level of conductance than the subgroup designated anxious (diff. = 7.2, σ diff. = 2.88, $t = 2.50$, $p = .02$).

DISCUSSION

Skin conductance, having been traditionally considered a physiological index of anxiety, is here viewed as a criterion against which discriminations made by the three dispositional indices of anxiety may be checked. The findings imply that the more specific an assessment technique is to the situation in which the predicted behavior is to occur, the more accurate is the prediction. This point is emphasized by the results obtained when S s were classified according to their standing on both the n -Ach and the TAQ measures. By this method, the nonsignificant between-subgroup trends observed in the individual analyses were further emphasized, reaching a significant level in two of the three experimental periods. Further, the difference in the amount of change between the two subgroups from the Anticipation to

the Solution Periods was accentuated as predicted.

An important question is raised by the present result: Of what importance are measures which are designated to be "general indicators" of anxiety? Behavior always occurs in specific situations. If an individual is termed "generally anxious," does it imply that he is anxious in all specific kinds of situations or just in a greater than average number? These results suggest that particular anxiety reactions are associated with specific situations. An assessment technique constructed to reflect response tendencies peculiar to a certain limited class of situations will probably have a predictive advantage over more generally oriented methods.

The analyses made in terms of the initial level of conductance in the Null 1 Period showed that in each of the individual difference measures, the subgroup considered to be the least anxious generally had the highest level of absolute conductance. Though these differences between subgroups in the individual analyses were not significant, they did reach statistical significance when *Ss* were classified on the combined *n*-Ach-TAQ indices—presumably the most reliable estimate of individual differences in disposition for achievement test anxiety. Since the Null 1 Period occurred before the experimental instructions were given, the differences must be due to initial reactions to the experimental situation itself, unless differences in conductance are a stable correlate of the individual difference measures.

Schlosberg (9) has argued that conductance levels should be considered an indicator of the degree of energy mobilization or "activation" of the individual in a particular situation. Activation has no particular "sign" or direction; it may stand either for positive interest in the task or for fear motivating avoidance. The *Ss* with a higher conductance level on entering the experimental room may be reacting to the situation with positive anticipatory interest cued off by the complex apparatus that immediately faces them. This interpretation is consistent with the contention of McClelland *et al.* (6) that the high *n*-Ach subgroup is likely to be challenged by a problem situation. Burdick (2) similarly reported a significantly greater tendency for persons high in *n* Ach to

volunteer to participate in psychological experiments.

If Schlosberg's theory of activation is correct, it may mean that the increased conductance level that the high *n*-Ach subgroup shows initially has quite a different "sign" than the change the low *n*-Ach subgroup shows later. The conductance measure itself is not capable of giving the direction of behavior implied by increased activation, but it is possible to infer direction from the way the two subgroups are conceptualized according to *n*-Ach theory. It is possible, therefore, to look at the conductance measure as an indicator of the degree to which an individual is activated by the cues of the situation. The direction of this activation, either positive (approach) or negative (avoidance), can be inferred through an analysis of the content of human thought (6, p. 33). Neither the MAS nor TAQ allow this possibility since there is no direct way of obtaining evidence of positive orientation from them.

SUMMARY

This study investigated relationships among three dispositional measures and two physiological indices of anxiety in a competitive achievement situation. The dispositional measures were relatively weak motivation to achieve as measured through content analysis of imaginative stories (low need Achievement), the Mandler-Sarason Test Anxiety Questionnaire, and the Taylor Manifest Anxiety Scale. The physiological indices were skin conductance and respiratory volume.

Twenty-five male college students were placed in a competitive test situation and instructed to perform a complex perceptual-motor task. They were informed when they made errors and were given cues to aid in performing the task accurately. The physiological indices were recorded before the ego-involving instructions were given and immediately before, during, and after the performance task. The three dispositional measures were obtained after the performance task.

The correlation between need Achievement and Test Anxiety was $-.43$; between Test Anxiety and Manifest Anxiety, $.53$; between Achievement and Manifest Anxiety, $-.25$. The two measures specifically concerned with reactions in competitive achievement situations (*n* Achievement, Test Anxiety), were both related to changes in skin conductance

during the performance task. The relationship was clearest when the subjects were classified on both of these measures as anxious (i.e., high Test Anxiety and low n Achievement) or nonanxious (low Test Anxiety and high n Achievement). The more anxious group increased in conductance while the relatively nonanxious group decreased. The Manifest Anxiety Scale did not relate to conductance change and no consistent relationships were found between respiratory activity and any of the other measures. These results suggest the advantage of situational over general measures of anxiety in accounting for changes in skin conductance in a particular situation.

It was also found that the highest initial level of conductance was obtained by the subgroup designated as nonanxious on each of the three dispositional measures respectively. These differences occurred before the introduction of the anxiety-arousing instructions. These findings are interpreted in terms of Schlosberg's theory of activation level.

REFERENCES

1. ATKINSON, J. W. Exploration using imaginative thought to assess the strength of human motives. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. of Nebraska Press, 1954.
2. BURDICK, H. A. The relationship of attraction, need Achievement and certainty to conformity under conditions of simulated group atmosphere. Unpublished doctor's dissertation, Univer. of Michigan, 1956.
3. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120-125.
4. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
5. MCCLELLAND, D. C. *Personality*. New York: Sloane, 1951.
6. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
7. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166-173.
8. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. C. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561-565.
9. SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81-88.
10. SNEDECOR, G. W. *Statistical methods*. Ames: Iowa State Press, 1946.
11. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. psychol.*, 1953, **45**, 116-119.
12. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.
13. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
14. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
15. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.

Received February 6, 1956.

CONDITIONING OF A RESPONSE CLASS ON A PERSONALITY TEST

ANNE M. NUTHMANN¹

University of Wisconsin

THE demonstration of verbal learning without awareness of what is being learned has been the subject of much controversy in psychological literature for many years. Recent studies (7, 8, 15) using a word-association technique report conflicting results, but positive findings (7, 8) are of very small absolute magnitude. Studies using other methods (1, 2, 3, 4, 12, 14) that might be expected to involve less interference from the subject's (*S*'s) previous experience have found more striking evidence for learning without awareness. In some (1, 2, 3, 14), verbal stimuli contingent upon *S*'s response have been found to have reinforcement properties, but when mechanical stimuli such as a light or a tone are used, reinforcement sometimes does (4, 12) and sometimes does not (1, 14) occur.

If learning can occur without awareness in areas relevant to personality formation when contingent stimuli are presented in the relatively simple fashion of these studies, the implications for psychotherapeutic methodology are great. A personality variable that has been given much attention in recent literature on psychotherapy is that of "attitudes toward self." Changes from more negative to more positive statements about the self have been shown to occur in successful therapy (9, 10, 11). Also scores on several scales of the MMPI have been shown to change significantly after therapy (5). There has been much less success in correlating therapist behavior with therapeutic outcome (13). The present study is an attempt to demonstrate an increase in responses indicative of acceptance of self on a personality test as a function of the contingent stimuli "good" and a light blink.

METHOD

Selection of items for the personality test. Several hundred items, many from existing personality and interest inventories, the remainder devised by the author, were classified by 80 students of elementary psychology courses at the University of Wisconsin into three categories: acceptance of self, rejection of self, and items having nothing to do with acceptance or rejection of self. Acceptance and rejection of self were

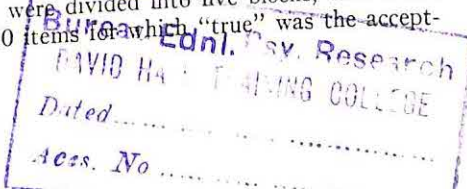
defined for the students as follows: "One accepts himself by thinking he is an adequate person, admitting that he has strong points and weaknesses. One rejects himself by thinking he is an inadequate person or by refusing to admit weaknesses which are true of almost everybody." These definitions were written at the top of the list of items which the students were required to classify, but no example items of the three categories were given. An item was included in the experimental test as indicating acceptance or rejection of self if the difference between the proportion of students placing it in one of these categories and the proportion expected by chance was significant at the .001 level. To meet this criterion, it was necessary for 57 out of the 80 students to place it in the same category. The two categories, "acceptance" and "rejection," were then combined since an answer of "false" to a rejection item indicates acceptance. To reduce the probability of the experimental *S*s becoming aware of the conditioning process, buffer items from the third category were added.

The resulting test was administered to a new group of 80 students to determine the relative number of acceptance-of-self answers to each item and to eliminate items that were not answered in the opposite direction from the majority by at least 7 of the 80 students, since these items would be unlikely to be conditioned. The number of "true" and "false" answers indicating acceptance of self were then made equal. The final test consisted of 100 acceptance-of-self items and 40 buffer items.

Selection of subjects. The test was then administered to still another group of 420 students. The *S*s for the experiment were chosen from students scoring in the lowest 15 per cent of the acceptance-of-self distribution. They were matched on the basis of these scores and assigned to one of the three experimental groups, "good," light, or control. There were 15 *S*s in each group. Between one and two months after this pretest the experiment was conducted.

Apparatus. In the small room used for the experiment, the principal piece of apparatus was a 22- × 24-inch fiberboard partition painted flat black and mounted vertically on a table so that it separated *S* and *E*. A 2½- × 4¼-inch opening in the center of the partition served as a display for 3- × 5-inch white index cards containing the items. Immediately above the opening was a hole 1 inch in diameter covered with white paper, which could be illuminated by a 6-v. bulb on *E*'s side of the partition to produce an evenly diffused signal-light to *S*. Lever-type microswitches labeled T (true) and F (false) by means of white paint on the partition and were placed 2 inches above the base of the partition and 8 inches in from either side. The switches were wired to 6-v. bulbs on *E*'s side of the partition, and signaled *E* which microswitch had been depressed. The light signaling *S* was controlled by a silent switch that could be set by *E* to go on when the T or F switch was depressed, or not to go on for either. The items, typed on the index cards, were divided into five blocks, each block containing 10 items for which "true" was the accept-

¹ Now at Mental Hygiene Clinic, Great Falls, Montana.



ance-of-self response, 10 items for which "false" was the acceptance-of-self response, and 8 buffer items. The blocks were matched for the number of acceptance-of-self answers to the items within them given by the group of 80 students to whom the test was first administered as a personality test. The order of presentation of the blocks was varied according to a 5×5 latin square to control further for possible unequal conditionability of the blocks.

Conditioning. In the experimental session *S* was seated at the table facing the fiberboard partition and was read the following instructions:

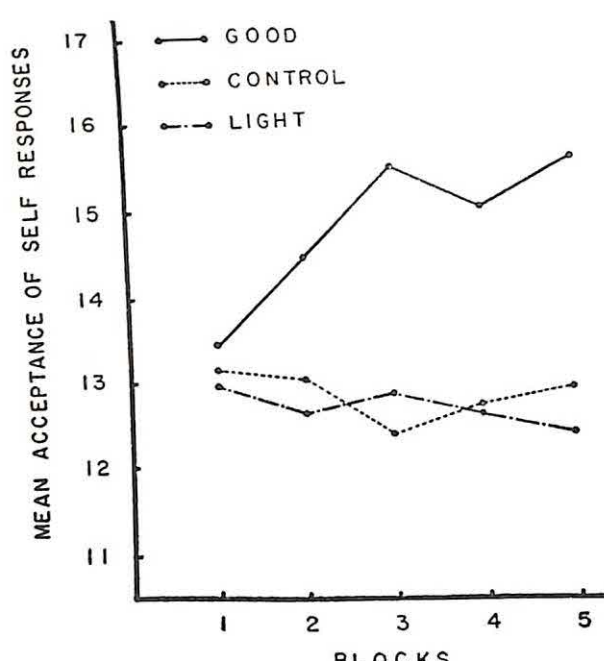


FIG. 1. MEAN ACCEPTANCE-OF-SELF RESPONSES FOR THE THREE GROUPS ACROSS THE FIVE BLOCKS OF ITEMS

TABLE 1
SUMMARY OF ALEXANDER TREND ANALYSIS
OF THE NUMBER OF ACCEPTANCE-OF-
SELF RESPONSES FOR THE THREE
GROUPS

Source of Variation	Error Term	df	Mean Square	F
a. Between group means	b	2	101.07	3.52*
b. Between individual means	h	42	28.75	8.03***
c. Group deviations from estimation	h	6	.99	—
d. Between group slopes	e	2	20.01	4.03*
e. Between individual slopes	h	42	4.97	1.39
f. Over-all slope	e	1	4.11	—
g. Over-all deviations from estimation	h	3	.47	—
h. Individual deviations from estimation		126	3.58	

* $p < .05$.

*** $p < .001$.

"Several weeks ago in your Psychology 1 class you took a personality test. I am interested in the effect that different ways of presenting personality statements has on the way people reply to them. You will now be given the same test you had before only in somewhat different form. Through this window you will be shown a series of cards on which are typed the personality statements. You are to read each statement carefully and decide whether or not it is true or mostly true with respect to yourself as you feel right at this moment. Then press the appropriate true or false switch in front of you. Hold the switch down for about one second, like this (*E* demonstrates), but do not press a switch unless a card is showing in the window. Try not to take too long to decide on your answers and don't worry about being consistent. It is a very common occurrence for people to change their answers on personality tests when they take them a second time. I am interested in your first impressions right now. I am not keeping a record of your name as I am interested only in group results, so please be as honest as you can in your responses. Are there any questions?"

The *E* then began to present the cards, the rate of presentation being determined by *S*'s response rate. For the two groups to be conditioned, whenever *S* gave an acceptance-of-self response it was followed immediately by the appropriate contingent stimulus. For all groups *E* then removed the card, recorded the acceptance-of-self response, and presented the next card. For the light group *E* set the light-control switch immediately before presenting each card.

After all 140 cards had been presented, *S*s in the light group were asked the following question, "What did you think the purpose of the light was?" The *S*s in the "good" group were asked, "Did you think that anything I said had anything to do with your responses?" On the basis of *S*s' answers to these questions *E* determined awareness of the conditioning process.

RESULTS

The principal findings of the experiment are shown in Fig. 1, in which the mean number of acceptance-of-self responses for the five successive blocks of items is plotted for each of the groups. The distributions within the various subgroups were moderately skewed, and when the data were subjected to Bartlett's test for homogeneity of variance of the *S*s within groups, a chi square of 7.52 was obtained. For 2 *df* this is significant between the .05 and .02 levels, indicating moderate heterogeneity of variance. Lindquist (6) suggests that nonnormality and heterogeneity of variance do not have an appreciable effect on the validity of the *F* ratio unless these deviations are marked. Therefore an Alexander trend analysis was carried out on the data of the three groups after a Bartlett's test of the slopes failed to show heterogeneity of variance (chi square of 3.29 for 2 *df*). The results of the trend analysis are summarized in Table 1. The

F of 3.52 for Between Group Means indicates that the means of the three groups differed to an extent significantly greater than could be accounted for by individual differences. The F of 4.03 for Between Group Slopes was also significant beyond the .05 level. Inspection of Fig. 1 indicates that both of these differences are almost entirely due to the progressive increase in acceptance-of-self responses for the group receiving the contingent stimulus "good." The curves for the light and the control groups are very similar over all five blocks, indicating that the light was not a reinforcing stimulus.

With respect to the performance level of the three groups at the end of Block 1, it is interesting to note that although the groups were matched on the basis of performance on the pretest before the beginning of the experiment it was not possible to match them exactly. The average score for Ss in the control group on the first test was slightly higher than that of the other two groups. The "good" group overcomes this handicap in the first block of items, suggesting that conditioning takes place from the beginning. However, though it is evident that the "good" group as a whole increased in acceptance-of-self responses over the five blocks, not all of the Ss in this group did so.

A latin-square analysis of the data was also carried out to determine whether the individual blocks differed significantly from each other. The F of 5.98 for blocks was significant at the .001 level, and indicates that certain blocks of items were easier to condition than others, in spite of the fact that the items were originally assigned to the blocks on the basis of equal popularity.

In the "good" group five Ss, and in the light group four Ss, became aware of, i.e., could verbalize, the purpose of the experiment. It is fairly certain that those who could not verbalize the correct purpose actually were unaware, because for the most part these Ss verbalized a contradictory hypothesis, many of them stating that reinforcement occurred whenever their response was consistent with that of previous testing. Since no over-all increase occurred for the light group, only the results of the "good" group were analyzed further. An Alexander trend analysis of the data for the aware and unaware Ss of the "good" group showed that neither the difference be-

tween the group means nor the difference between the group slopes was significant, the F ratio for the slope difference being less than 1.00. The aware group, however, did respond at a slightly higher level than the unaware group at all points.

Since the Ss for this experiment represented one extreme of a distribution, one might expect a regression toward the mean to occur for all the groups on a second testing. There actually was an over-all average increase in total acceptance-of-self responses from the pretest to the experimental test. In the control group this average total increase was 6.74 responses, in the light group 6.53 responses, and in the "good" group 16.86 responses. In the "good" group 7 of the Ss increased 20 responses or more; 11 of them increased 10 responses or more.

DISCUSSION

The results of the present experiment demonstrate that it is possible to learn without awareness, defined in terms of verbal report, to respond in a more self-accepting way to items on a personality test. Here, as in other experiments on learning without awareness, one may ask exactly what it is that the Ss learned. In the studies (3, 4, 12) where plural words were conditioned, for example, did the Ss learn the concept plurals or did they learn to say words ending with the *s* sound? In the present case it is obvious that the Ss did not learn to respond in a certain way to specific items, because each item occurred only once. They did not learn a "true-false" pattern of responding because the items to be conditioned did not appear in such a systematic order. Also a given order of presentation of the blocks cannot account for the group increase in acceptance-of-self responses since this order of presentation was varied in a latin square design. Another possibility arises as to the effect of differential wording of the items to which "true" and "false" were the acceptance-of-self responses. Very few of the rejection-of-self items were worded negatively, so that the Ss could not learn to respond "false" to negatively worded items. It is true, however, that most of the times for which "false" was the acceptance-of-self response did contain words that might be classified as being negative in emotional tone, whereas the "true" acceptance-of-self items in general contained words

1-9-76

of positive emotional tone. Whether emotional tone or the broader concept of acceptance of self was actually discriminated remains open to question. In favor of the latter alternative is the consideration that if discrimination of emotional tone did occur, some of the Ss should have been able to formulate it verbally, since some of them were able to verbalize the more abstract concept of acceptance of self. For the purpose of this discussion, therefore, it is assumed that discrimination was in terms of self-acceptance.

The verbal stimulus "good" was shown to be capable of bringing about this increase in acceptance-of-self responses. There were, however, large individual differences in the amount of conditioning. Two factors may be suggested to account for these differences. As a result of a differential degree of past learning by the Ss of responses involving their self concept, it may take longer for reinforcing stimuli to bring about a measurable change in some Ss' behavior than in the behavior of others. That is, Ss who have responded, overtly or covertly, "I am worthless" hundreds of times may take longer to change this response than Ss who have thus responded only a few times. Also the stimulus "good" may have different reinforcing value for different Ss as a result of their previous experiences with it.

The behavior of Ss in the group who were aware of the *E*'s intent, it has been noted, did not differ significantly from the behavior of those who were unaware. Four of the five "aware" Ss volunteered the additional information that they were quite sure that they had not been influenced by the reinforcement.

The results obtained with the verbal stimulus "good" may have important implications for psychotherapy. It is possible that some of the changes in a patient's verbalized self concept that occur in therapy are the result of simple verbal reinforcement given, perhaps sometimes unknowingly, by the therapist.

The light blink, however, clearly was not a reinforcing stimulus in this situation, not even for the four Ss who were aware of the purpose of the light. This finding offers evidence to support a previously suggested hypothesis (14) that a situation of the clinical sort is not conducive to conditioning by means of a light. Perhaps in order to function as a reinforcer of verbal behavior a stimulus must in some way be familiar to the *S* as having a connection

with his responses. In the studies that found the light to be reinforcing (4, 12) Ss were not responding with meaningful verbal material, but were simply saying words. Presumably the Ss had never experienced this type of situation previously, and therefore a blinking light would not be a strange resultant of their behavior. On the other hand, in the studies obtaining negative results with a light (1, 14) and in the present study, the Ss certainly had much previous experience with using meaningful sentences but most probably had no experience with being confronted with a light blink as a conveyor of feedback.

SUMMARY

The purpose of the present experiment was to ascertain the reinforcement value of the verbal stimulus "good" and a light blink stimulus on a class of behavior designated as acceptance-of-self responses on a true-false personality test. Each item in this test that was to be conditioned was chosen on the basis of agreement among student judges significantly different from chance at the .001 level that the item measured acceptance of self. The test was then administered to 420 psychology students, and Ss for the experiment were chosen from a subpopulation receiving low acceptance-of-self scores.

The Ss were assigned to one of two experimental groups or a control group after having been matched on the basis of their performance on the pretest. In the experimental situation the items were presented to the Ss individually on index cards. For Ss in the experimental groups, each true or false answer indicating acceptance of self was reinforced by *E*'s saying "good" (in one group) or by an immediate light blink (in the other.) In the control group no feedback of any kind was given. The S's awareness of the conditioning procedure was determined from his verbal report at the end of the experimental session.

The results indicate that (a) learning without awareness to respond positively to acceptance-of-self statements can occur; (b) the behavior of Ss who are aware of the purpose of reinforcement does not differ significantly in this situation from that of Ss who are not aware; (c) the verbal stimulus "good" is capable of bringing about such learning whereas the nonverbal light stimulus is not.

REFERENCES

1. BALL, R. S. Reinforcement conditioning of verbal behavior by verbal and non-verbal stimuli. Unpublished doctor's dissertation, Univer. of Indiana, 1952.
2. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106-110.
3. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409-416.
4. GREENSPOON, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctor's dissertation, Univer. of Indiana, 1950.
5. KAUFMANN, P. Changes in the Minnesota Multiphasic Personality Inventory as a function of psychiatric therapy. *J. consult. Psychol.*, 1950, **14**, 458-464.
6. LINDQUIST, E. P. *Design and analysis of experiments in psychology and education*. Cambridge: Riverside Press, 1953.
7. PHILBRICK, E., & POSTMAN, L. A further analysis of "learning without awareness." *Amer. J. Psychol.*, 1955, **68**, 417-424.
8. POSTMAN, L., & JARRETT, R. F. An experimental analysis of "learning without awareness." *Amer. J. Psychol.*, 1952, **65**, 244-255.
9. RAIMY, V. Self-reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153-163.
10. SEEMAN, J. A study of the process of nondirective therapy. *J. consult. Psychol.*, 1949, **13**, 157-168.
11. SHEERER, E. T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169-175.
12. SIDOWSKI, J. B. Influence of awareness of reinforcement on verbal conditioning. *J. exp. Psychol.*, 1954, **48**, 355-360.
13. SNYDER, W. U. Supplementary correlations relative to the relationship between the use of client-centered techniques and the success of psychotherapy. In *Group report of a program of research in psychotherapy*. State College: Pennsylvania State Col., 1953.
14. TAFFEL, C. Conditioning of verbal behavior in an institutionalized population and its relation to "anxiety level." Unpublished doctor's dissertation, Univer. of Indiana, 1952.
15. WEISS, R. L. The influence of "set for speed" on "learning without awareness." *Amer. J. Psychol.*, 1955, **68**, 425-431.

Received February 14, 1956.

THE GENERALITY OF ATTITUDES TOWARD AUTHORITY AND NONAUTHORITY FIGURES^{1, 2}

LEROY S. BURWEN

Research Division, Chicago Tribune

AND DONALD T. CAMPBELL

Northwestern University

THE collection of men into large-scale coordinated organizations typically involves a hierarchical order of authority in which the great bulk of persons are placed between a superior and a subordinate echelon. This common situation gives general importance to the description and diagnosis of any generalized attitude toward authority, if indeed such exists. Furthermore this is one predictive setting in which all available psychological theories seem to agree. Writers with as varied orientations as Freud, Piaget, Rogers, Newcomb, Krech and Crutchfield, etc., all agree in positing attitudes towards parental authority originating in the family situation which generalize and manifest themselves as attitudes toward superiors in later social situations. Piaget, for example, writes as follows:

Day to day observation and psychoanalytic experience show that the first personal schemes are afterward generalized and applied to many people. According as the first inter-individual experiences of the child who is just learning to speak are connected with a father who is understanding or dominating, loving or cruel, etc., the child will tend (even throughout life if these relationships have influenced his whole youth) to assimilate all other individuals to this father scheme (7, p. 207).

There is also a widespread awareness among clinical psychologists that problems of sibling rivalry repeat themselves in peer relationships in adult life. Whether these attitudes or response generalizations are mediated by stimulus similarity in terms of age and physical appearance; by language habits in which the terms father, old man, and boss are used interchangeably; or by learned stimulus equiv-

alence, role similarity, or other nonspatial configurational attributes, the predictions remain the same. The present study attempts to put such predictions to test.

In terms of methodological orientation, this study attempts to measure a small number of attitude dimensions through a large number of independent measurement techniques. The more common approach in which a large number of traits are measured by a single method leaves the investigator unable to distinguish between meaningful psychological relationships and such irrelevant sources of common variance as "apparatus factors" (9) and "response sets" (3, 4). The approach here taken is congruent with the notion of "construct validity" (8). All the measures introduced are accepted as subject to both random and systematic error. None is set up above the others as a "criterion." Validation becomes a process of mutual confirmation among a series of measures all directed toward the same psychological dimension, but maximally independent in terms of apparatus, conditions of administration, and the like.

METHOD

The data were collected during January and February of 1953 from 155 men assigned to 17 bomber crews at Randolph Air Force Base. The men were in crew assembly training in preparation for action in Korea and had been together for approximately two weeks. The 73 officers were typically men in their late 20's and early 30's with World War II experience. The 82 enlisted men were largely in the 19-23 age range in their first tour of military duty. Several crews were assembled at the same time for the group testing. The two hours of testing represented here was only a portion of the research testing to which they were subjected. In the testing there was emphasis upon the fact that the data were for research purposes and would not become a part of any man's record. The practical anonymity and mutual support of the group situation, plus more general attitudes characteristic of the impending transfer to Korea, resulted in a testing situation characterized in some cases by perfunctory compliance and occasional humorous sabotage of the test purpose. In the end, between 10 and 15 per cent of the answer sheets had to be discarded. In addition to

¹ This study was sponsored in part by the United States Air Force under contract No. AF 18 (600)-170 monitored by the Crew Research Laboratory, Air Force Personnel and Training Research Center, Randolph Air Force Base, Randolph Field, Texas.

² This article is in part based upon a dissertation submitted by the senior author to the University of Chicago in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

the group tests, 57 interviews were conducted individually on a basis of compulsory assignment. Particularly for the enlisted personnel, this setting created a guarded, deferential attitude that was difficult to overcome during the one-hour period involved.

Seven different measurement approaches were employed. Through each of these, as appropriate, attitudes were sought toward: own father; symbolic authority (as might be shown, for example, in responses to the picture of older persons in the TAT); immediate boss (the aircraft commander for most crew members); immediate peers (fellow crew members); and symbolic peers. In all, 21 separate variables were scored. Unless otherwise specified, the number of cases upon which the reliability coefficients are based is 155, except for the interview measures, where the N is 57. Officers and enlisted men have been pooled in the analyses that follow, in order to provide maximum stability for the coefficients computed. Significant officer-enlisted differences were found on only 3 of the 21 measures.

Procedures

1. *Interview.*³ The interviews focused upon family background and the attitudes toward the current Air Force situation. The subject (S) was drawn out on attitudes toward his father, siblings, present superior officers, and peers. Each of these four attitudinal areas was then scored on a five-point scale for favorability or unfavorability. These ratings were made on what the interviewer inferred the covert attitudes of the S to be, rather than simply the overt attitudes expressed. While no direct estimate of the reliability of interview ratings is available, a minimum estimate might be the squared correlation coefficient of the interview variable with other test variables (see Table 1).

2. *Thematic Apperception Test.* TAT-like pictures were administered by projection upon a screen while the S s wrote out their stories. The assignment to respondents was: "You are to tell a story about this picture which will explain what has happened and how things are going to turn out." Eight pictures were used, three from the Murray set (1, 6BM and 7BM) and five specially designed for Air Force use.⁴ Cards were selected so as to elicit attitudes toward Symbolic Authority and Symbolic Peers. Each of these attitude variables was scored on a five-point rating scale ranging from highly unfavorable characterization of the person to highly favorable characterization. Older persons and persons of higher military rank were assumed to be symbolic of authority figures while peer-age persons and persons of lower rank were interpreted as symbolic peers or subordinates. The interjudge reliability for these ratings is represented by a correlation coefficient of .83 between one judge and another on a sampling of 92 stories.

Internal-consistency reliability is difficult to estimate on a free-response test where not every person gives a scorable response to every stimulus. Seventy-one S s

gave some scorable Symbolic Peer descriptions on two of the cards. The correlation between the ratings received on these two cards was .36, which corrected by the Spearman-Brown formula yields a reliability of .53 for a total score based upon the two cards. Scorable responses for Symbolic Peer also were made to other cards. The mean number of scorable Symbolic Peer responses per S for the entire 155 S s was 1.8. One hundred and eight S s gave scorable responses to presumed authority figures on four of the cards. Correlating the sum of the ratings on two of these cards with the sum of the ratings on the other two yields an r of .38 which corrects to .55 for the reliability of the sum of the four card scores. The mean number of scorable Symbolic Authority responses for the total 155 persons was 6.2; thus the reliability based upon four cards is presumably a slight underestimate. The scoring procedure employed was somewhat more direct than many clinical TAT interpretations. In particular, no use was made of the presumed ego representation in the story, nor of more complicated dynamic interpretations (e.g., such as would involve reaction formation). It was therefore thought desirable to rate the stories on a more clinical basis, using whatever complex cues might be available. It is important to note here, however, that the stories contain much less complexity than those provided by neurotic persons with middle-class backgrounds. These global clinical ratings are designated by TATg while those based upon scoring the descriptions of specific persons summed over the eight pictures are indicated by TATO.

3. *Description of Self and Others.* In this test the respondents were asked to evaluate other persons by indicating whether each of 30 different trait terms applied or did not apply. Eight of the 30 terms were to be marked as applying; eight as definitely not applying; and the rest omitted with a resulting intermediate weighting. The descriptions of father, immediate superior (boss) and a fellow worker were scored in terms of the degree to which favorable terms were used in the description. The reliabilities⁵ are as follows: father, .24; boss, .34; peer, .55.

4. *Photo Judging Test.* In this test the respondents were asked to characterize briefly persons shown in photographs. It was assumed that the photographs of middle-aged and older persons would be most likely to tap attitudes toward those superior in authority, and that reactions to the photographs of high school age persons would be most likely to tap attitudes toward subordinates and peers. For the present scoring purposes the responses to 20 photographs of men over 45 form the basis for the score on attitudes toward Symbolic Authority. The responses to five photographs of high school boys provided the score labeled Sym-

⁵ Unless otherwise noted, the reliabilities referred to for this and following measures are internal consistency measures of the Kuder-Richardson type, with the computation using item variances and total score variances, following Gulliksen's Formula 10 (5, p. 223). If the assumptions underlying this coefficient are not met, the reliability is systematically underestimated. It is worthy of note that although the reliability of father description is only .24, it correlates .40 with both the interview rating on father and with the father score obtained from the autobiographical inventory.

³ The senior author, with a background of four years of clinical work, conducted the interviews and scored all measures involving clinical or subjective judgment.

⁴ The authors are indebted to Dr. William Soskin for two of these pictures; three were designed by the senior author.

TABLE 2
MEAN CORRELATIONS BY TOPIC

Topic		Topic				
		F	SA	B	P	SP
Father	F	.35	.12	.03	.06	.08
Symbolic Authority	SA		.15	.08	.10	.06
Boss	B			.09	.13	.03
Peer	P				.22	.07
Symbolic Peer	SP					.01

bollic Peer. The assignment to the Ss was to "Write down what you think the person in the picture is like." Space was left for 2 or 3 sentences, although in general a listing of trait terms was given. These responses were rated as to whether the characterization was favorable, intermediate, or unfavorable. Interjudge agreement in making these ratings was .95 for one pair of judges and .86 for another. Internal consistency reliability for Symbolic Authority was .46, for Symbolic Peer, .29.

5. *Autobiographical Inventory.* This inventory contained both short free-response items and check-list items. The full details of the questionnaire are too long for reporting here, and, as with all instruments discussed in this paper, have been reported fully elsewhere (2). Scoring involved both subjective ratings of free-response items and direct scoring of check-list material. Interjudge agreement on a sample of 100 subjectively evaluated items was .86. There were eight items on the S's relationship to his father and these yielded an internal consistency reliability of .56. Attitudes toward boss on this instrument included present boss and previous supervisors and teachers. There were 16 such items which yielded a reliability of .56. There were 11 items describing relationships with peers, including classmates, childhood friends, and fellow workers, giving a reliability of .55.

6. *Attitude Survey.* The tested personnel were also given an attitude survey developed for the study of morale by the Crew Training Research Laboratory, Randolph Air Force Base. Ten of the items of this test, all of which were statements to be endorsed in one of five levels of intensity, seemed to reflect attitude toward authority in the military situation. These items included competence of superior officers and instructors, and general favorableness of attitudes toward the Air Force as an administrative unit. Scoring the responses with five levels of weighting, the symbolic authority score from this instrument had a reliability of .59. Ten items reflected on attitudes toward members of the crew as peers and had a reliability of .73. The peer items were considerably more homogeneous topically than the Symbolic Authority items. These items represented attitudes toward actual peers in the working situation, while the authority items were of such a general nature that we have classified them as attitudes towards Symbolic Authority.

7. *Sociometric Questionnaire.* The "Position Description Test" developed by the Crew Training Research Laboratory required each crew member to pick those fellow crew members whom he would like to be with in five hypothetical situations, including a crew party, shared work responsibility, bailing out behind enemy lines, etc. Each person's attitude toward boss was

based upon the relative proportion of his mentions given to the aircraft commander. (Aircraft commanders as respondents were left out of this analysis, reducing the total number of cases to 138.) Peer scores for the officers were based upon the nominations given the other officers; for the enlisted personnel, the score was based upon the responses given to the other enlisted persons. For boss, the reliability was .75. For peer, the values were .84 for officers and .90 for enlisted men.

RESULTS

Table 1 presents the product-moment correlations among these 20 variables. The variables are arranged according to the a priori attitude groupings and the expected pattern of correlations. The a priori expectations were of generalized attitudes among the measures dealing with father, symbolic authority, and boss and also homogeneity of responses among the various measures focused on peers and symbolic peers. A relative independence between these two major sets of attitudes was expected.

The ubiquitousness of "apparatus factors" (9) is indicated by the fact that for 15 of the 20 variables the highest correlations were obtained between measures of different attitudes based upon the same instrument. For only 3 of the 20 variables is the highest correlation with another measure of the "same" attitude through a different instrument. This "apparatus factor" effect is particularly marked for the Interview, TAT, and Sociometric measure. In Table 1, correlations involving the same instrument have been put in parentheses so that they may be discounted in the interpretation. It is obvious that in general the other correlations are considerably lower than expected.

In spite of the disappointingly low general level of correlation, there is a slight general positive trend. Disregarding the intra-instrument coefficients, the values range from $-.28$ to $+.56$. Of the 169 coefficients not involving the same instrument, 143 are positive, differing from a chance distribution at the .001 level, as tested by the binomial expansion.

To examine this matrix for evidence of specific attitudes toward father, boss, peers, etc., Table 2 has been prepared from Table 1 by averaging by z transformation the correlations within each of the cells, eliminating those based on the same instrument, and weighting by number of cases (which reduces the influence of the interview measures). If specific attitudes exist, correlations among

TABLE 3

COMPARISON OF CORRELATION PAIRS TO TEST THE ASSUMPTION OF STIMULUS GENERALIZATION FROM FATHER TO BOSS

Father Measure	Corr. With	Measure of Boss and Peer Attitudes				Score
		Int.	S & O	A.I.	Soc.	
Int. Father	Boss	.64	.19	-.06	.19	$\frac{1}{4}$
	Peer	.65	.27	.04	.04	
S & O Father	Boss	.08	.23	-.14	.06	$\frac{1}{4}$
	Peer	.09	.21	.00	.08	
A. I. Father	Boss	.15	.09	-.08	-.03	$\frac{1}{4}$
	Peer	.24	.03	.15	.05	
						$\frac{3}{12}$

variables focused upon the same attitude object should be higher than correlations among measures involving dissimilar attitude objects. The data in Tables 1 and 2 indicate that this requirement is clearly met only for attitudes toward father and toward peers. It is most clearly not met for symbolic peers and boss. Even where met, the level of inter-correlation is much lower than one might have expected.

In attempting to interpret these findings it is well to look more closely at the assumptions involved in giving the different measures similar labels. In anticipating some equivalence among the various measures of attitude toward father we assume *response generalization* or response equivalence to a relatively constant, identifiable stimulus. The same assumption is involved for the measures of attitude toward boss and peer, with the exception of the scores taken from the autobiographical inventory, in which the measure is based upon responses to a number of previous bosses and previous peers. Removing the autobiographical inventory measures for boss and peer should thus improve the extent to which they refer to a single attitude object. So doing raises the average correlation among the measures of peer attitude about .05 points and has a similar effect upon the correlations among measures of attitude toward boss. However, even with this alteration, measures of attitude toward boss still fail to meet the simple requirement that the measures correlate higher among themselves than with measures focused on other attitude objects;

even with all autobiographical inventory measures removed, measures of attitude toward boss correlate more highly with measures of attitude toward peer than they do with other measures of attitude toward boss.

The other assumption employed in the labeling of the attitude measures is that of *stimulus generalization*. It is on the basis of assumed stimulus generalization that the symbolic peer attitude measures were given similar labels, and that the symbolic authority attitudes were placed in the same category. On the basis of one of the most widespread hypotheses in the psychology of personality we also assumed generalization from father to symbolic authority and to boss (1, 6, 7). In Table 2, it is clear that the three presumed groups with respect to attitude toward authority (father, symbolic authority and boss) correlate no more highly with each other than they do with the two sets of peer measures. Thus while we did find some evidence of specific attitudes toward father, we find no evidence at all for a generalized "attitude toward authority" which would include in its manifestations attitudes both toward father and toward persons in supervisory positions. This finding is so out of keeping with general theoretical anticipations that it is worth examining in greater detail.

The comparisons just made are somewhat unsatisfactory because of the variability or heterogeneity of the correlation coefficients within the blocks averaged. Not all of the attitudes have been measured with all of the instruments, and the instruments differ in the strength of apparatus factors. In addition, stimulus generalization or habit similarity may be expected on a priori grounds, involving social objects sharing the same temporal and spatial situation, and leading in particular to factors centering around the early home, on the one hand, and the present working situation on the other. The correlation between peer and boss attitudes might, for example, be due to the fact that they are both specific manifestations of a more generalized attitude toward the current working situation. To achieve comparisons in which these factors are held constant, we have extracted from Table 1 specific pairs of correlations that are comparable with regard to the instrument and the situational factors involved. Some of these correlation pairs are shown in Table 3. All

TABLE 1
INTERCORRELATIONS OF ATTITUDE MEASURES

Topic and measure		Topic and Measure																			
		Father			Symbolic Authority					Boss				Peer					Symbolic Peer		
		Int.	S & O	A.I.	Int.	TATg	TATo	Photo	Surv.	Int.	S & O	A.I.	Soc.	Int.	S & O	A.I.	Surv.	Soc.	TATg	TATo	Photo
F	Interview	.40	.09	(.83)*	.14	.08	-.08	.41	(.64)	.19	-.06	.19	(.65)	.27	.04	.07	.04	.35	.30	-.15	
	Description of Self and Others				.17	.18	.09	.14	.08	(.23)	-.14	.06	.09	(.21)	.00	.05	.08	.11	-.03	.09	
	Autobiographical Inventory		.40		.07	.19	.07	-.08	.13	.15	.09	(-.08)	-.03	.24	.03	(.15)	.05	.05	.08	.12	.07
SA	Interview				.19	.17	-.04	.51	(.81)	.12	-.14	.37	(.77)	.23	.04	.20	.16	.36	.20	-.05	
	TAT (scored globally)					(.72)	.13	.21	.14	.13	.03	.04	-.05	.08	.12	.16	.01	(.49)	(.30)	.07	
	TAT (scored objectively)						.12	.22	.23	.07	.00	.07	.03	.07	.04	.15	.01	(.32)	(.19)	.01	
	Photo Judging Test							.00	-.12	.07	.01	.01	-.28	.14	.04	.11	.02	.14	.00	(.15)	
	Attitude Survey #6								.56	.14	.03	.15	.26	.30	.09	(.41)	.11	.11	.03	.00	
B	Interview									-.10	-.17	.40	(.76)	.11	.09	.18	.26	.17	.14	-.02	
	Description of Self and Others										.09	.16	-.03	(.45)	.04	.25	.10	.10	-.02	.05	
	Autobiographical Inventory											.02	-.20	.03	(.15)	.11	-.04	-.09	.01	.04	
P	Sociometric												.05	.41	.17	.20	(.58)	.06	-.03	.06	
	Interview													.23	-.04	.05	.08	.23	.08	.20	
	Description of Self and Others														.20	.31	.35	.01	-.05	.17	
	Autobiographical Inventory															.19	.05	.05	-.03	.20	
	Attitude Survey # 6																.36	.05	.04	.15	
SP	Sociometric																	.05	.01	.15	
	TAT (scored globally)																		(.65)	.06	
	TAT (scored objectively)																			-.04	
	Photo Judging Test																				

* Parentheses indicate correlations between attitude measures from the same instrument.

of the pairs of coefficients in this table represent correlations of some one measure of attitude toward father with measures of boss and peer attitudes from the same instrument. The boss and peer attitude measures are comparable, inasmuch as they represent the same environmental factor and are diagnosed by the same instrument. They differ only in their predicted similarity to father. As indicated in Table 3, interview attitude toward father correlates .64 with interview attitude toward boss, and .65 with interview attitude toward peers. Interview attitude toward father correlates .19 with Description of Self and Others attitude toward boss, and .27 with Description of Self and Others attitude toward peer, etc. Each of the pairs is comparable in terms of the extraneous factors of instrument type and situational sources of variance. If there is stimulus equivalence between father and boss, the correlation with boss within every pair should be higher than the correlation with peer. Similar sets are presented for each of the three approaches to measuring attitudes toward father. The column headed "score" indicates the proportion of instances in which the expected relationship was obtained—in only three pairs out of the twelve, over all. Thus, not only do these carefully selected comparisons fail to support the existence of stimulus generalization from father to boss but also they show a trend approaching statistical significance in the opposite direction.

In a manner similar to the analysis in Table 3, a number of other comparisons are possible. A higher correlation may be expected between attitudes toward father and attitudes toward symbolic authority than between attitudes toward father and attitudes toward symbolic peers. The TAT scored globally, the TAT scored objectively, and the Photo Judging Test all yield measures of attitudes toward symbolic authority and symbolic peers. If correlations between these two variables and the three measures of attitudes toward father used in Table 3 are compared, a successful prediction is achieved in only 4 out of 9 cases.

Through analogous procedures the following additional stimulus generalization hypotheses were tested, with the outcomes indicated: symbolic authority and boss should correlate higher than symbolic authority and

peer (10/20); boss and symbolic authority should correlate higher than boss and symbolic peer (7/12); peer and symbolic peer should correlate higher than peer and symbolic authority (8/15); symbolic peer and peer should correlate higher than symbolic peer and boss (7/12). Needless to say, none of the outcomes approaches statistical significance.

DISCUSSION

At the inception of this study it was anticipated that persons bring with them into any social situation a repertoire of attitudes or habits which modify their interaction with the persons they meet. The literature of social psychology and personality supports the expectation that prominent within this repertoire of habits or generalized attitudes are attitudes toward authority derived from previous encounters with authority, especially in the early family situation. Basic to this kind of anticipation is an assumption of a symbolic similarity between father and other kinds of authority and a generalization of attitudes from one to the other. With respect to the existence of any such generalized attitude toward authority, independent of a still more general attitude toward people, or a still more specific attitude toward specific father or specific boss, this study has been entirely negative. Furthermore, it has shown no evidence of stimulus equivalence between symbolically represented authority figures and actual authority figures or symbolically represented peer figures and actual peer figures. These findings are so out of keeping with the general assumptions of current social and personality psychology of all schools that they are worth some special consideration.

The simplest interpretive device in situations such as this is to reject the appropriateness of the instruments used. When the instruments are employed, one at a time, this approach is usually convincing, for in the application of any particular instrument there are many reasons why it is not wholly appropriate to the task at hand. In a multi-instrument approach, such as essayed here, it was certainly anticipated that some tests would prove better than others. However, the general level of confirmation of the hypotheses was so low that discriminations among the tests as to their efficacy or appropriateness

cannot be made. If tests are to be rejected, they must be rejected collectively. While the writers cannot claim any special excellence for the instruments employed, they do feel that the selection represents a considerable range of approaches unlikely to share common faults, and that collectively the battery is probably representative of the types of approaches that students of personality or social psychology would be likely to assume to be relevant for the problem studied. Rejecting the instruments is appropriate only if one is prepared to suggest a number of other instruments that would be superior and to specify the kind of application in which one would anticipate confirmation. Also, a rejection of the instruments is in itself a modification of the theory, inasmuch as it restricts or limits the kind of situations in which the theory is expected to be applicable. There may be stronger grounds for questioning the adequacy of the conditions of test administration, although the moderate obtained reliabilities are evidence against this position. But certainly parallel studies under more desirable testing conditions would be advantageous.

If one does not reject the results on these two grounds, then must one reject the theory? It is unlikely that such a recommendation would have any great impact upon current social psychological thinking, so it is probably better to speak in terms of modification of the theory in the direction of restricting its applicability. A sensible interpretation might be that although this kind of symbolic equivalence or stimulus generalization is appropriate to the autistic thought processes in normal dream states and in the waking states of extreme neurotics and psychotics, it does not interfere with the waking perceptions of normal individuals such as constituted the test population for the present study. This seems the most reasonable interpretation to the investigators, even though one would have expected some slight carry-over of autistic thought tendencies into our test situations.

The original predictions of this study seem consistent with the principle of stimulus generalization in conditioning theory. From Gestalt theory one would also expect generalization based upon configurational similarity from

one authority situation to another. Within the framework of Freudian theory, our results might be considered less surprising. Freud posits the similarity of opposites, and if reaction-formation mechanisms were present in our subjects, attitudes toward boss and father might conceivably be reversed in these instances. It seems to the writers, however, that in usual psychoanalytic presentations, our original expectations are regularly entertained. Even though our simple measure of favorableness toward father may be eliciting expression of reaction formations rather than basic attitudes, one would generally anticipate that the same reaction formation would be used by the person to conceal from himself and others his underlying hostility toward his boss; or that the same protestations of rebellion which the boy manifested toward his father would be repeated toward the boss, to cover, for example, persistent underlying dependency needs. To explore this possibility, we have searched the correlation matrix for possible evidence of greater agreement from depth measure to depth measure, and from surface measure to surface measure, than between a depth and a surface measure. No such evidence is to be found. Furthermore, the scatter diagrams for each of the coefficients in Table 1 have been inspected without success for curvilinearity or more complicated patterns that concepts such as reaction formation would predict. Only in a complex analysis of inter- and intra-individual variability have evidences of dynamic personality processes been found. These analyses will be reported later and do not reduce the incongruity between our present findings and common assumptions in the area of personality and interpersonal attitudes.

SUMMARY

Presumed attitudes toward father, symbolic authority, boss, peers and symbolic peers were measured by interview, descriptive trait check lists, autobiographical inventory, thematic apperception test, judgment of photographs, an attitude survey and a sociometric questionnaire. Intercorrelations among the 21 measures so obtained are generally low, except where two or more measures come from the same instrument. In such in-

stances the high correlations seem to reflect apparatus factors rather than psychologically meaningful communalities. The total matrix of correlation is generally positive to a significant degree. Of the five attitude topics, only among attitudes toward father and among attitudes toward peer are the correlations of such a nature as to indicate that a specific attitude focus is to some extent being tapped by independent methods. Even here the degree of confirmation is not high. For attitudes toward father, the average correlation among three separate measures is .35; for attitude toward peers, the average correlation among five independent measures is .22. Evidence for a generalized attitude toward authority which encompasses attitudes toward father, symbolic authority, and boss is totally negative, suggesting the need for reconsideration of the applicability of commonly held theory in this area.

REFERENCES

1. ALLPORT, G. W. Attitudes. In C. Murchison (Ed.), *Handbook of social psychology*. Worcester: Clark Univer. Press, 1935. Pp. 798-844.
2. BURWEN, L. S. A study of attitudes towards authority. Unpublished doctor's dissertation, Univer. of Chicago, 1954.
3. CRONBACH, L. J. Response sets and test validity. *Educ. Psychol. Measmt*, 1946, **6**, 475-494.
4. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3-31.
5. GULLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.
6. MURPHY, G., MURPHY, LOIS B., & NEWCOMB, T. M. *Experimental social psychology*. (Rev. Ed.) New York: Harper, 1937.
7. PIAGET, J. *Play, dreams and imitation in childhood*. New York: Norton, 1951.
8. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.*, 1954, **51**, (Suppl. No. 2, Pt 2) 1-38.
9. TRYON, R. C. Individual differences. In F. A. Moss (Ed.), *Comparative psychology*. (2nd Ed.) New York: Prentice-Hall, 1942. Pp. 330-365.

Received February 16, 1956

GUESSING BEHAVIOR AND AUTISM¹

CHARLES M. SOLLEY, DOUGLAS N. JACKSON,
The Menninger Foundation The Pennsylvania State University
AND SAMUEL J. MESSICK
Educational Testing Service

WHENEVER individuals are required to guess the occurrence of alternative future events without any specific knowledge of what to expect, they probably rely upon previously established expectancies. As *S* gains more and more experience with these "future" events, he can modify his guessing behavior to accord with the "true" probability of occurrence of such events. The present investigators were interested in *Ss'* initial guesses in a probability learning task where motivational factors might influence *Ss'* guessing behavior. The primary hypothesis was that *Ss* with minimal information would overguess the frequency of occurrence of stimulus events when those events had been previously associated with signs of goal attainment or need-satisfaction. A secondary purpose of these studies was to investigate some semantic aspects of Osgood's mediation theory (9).

The primary hypothesis was derived initially in the following way. First let us assume that some stimulus event has been directly associated with goal-reaching behavior or need-satisfaction; i.e., it has become a sign for need-satisfaction and concepts related to it. Further let us assume that stimulus events that signify goal-reaching and need-satisfaction tend to be perceived as larger, brighter, denser, or more numerous than they really are (1, 2, 3, 4, 5, 6, 13). That is, we are dealing with classes of stimuli that are associated with need-satisfaction, and we consider signs of need-satisfaction to be accentuated in perception. Now, if neutral stimuli are associated with these signs, they might also acquire the

same perceptual properties as the signs, and be perceived as larger, brighter, denser, or more numerous than they really are. It would seem to follow, then, that if a neutral stimulus is thus perceived as more numerous, for example, *S* may overguess the frequency of its occurrence in probability learning situations.

In the present study certain stimulus events were selected that could be assumed to be related to need-satisfaction or positive value. In order to choose such stimulus events, reference was made to factor analytic studies of meaning reported by Osgood and Suci (10). These studies were based upon correlations among several bipolar scales, the ends of which are defined by adjectives such as *good-bad*, *strong-weak*, etc. Of the three factors extracted—evaluation, potency, and activity—the evaluative factor appeared to reflect meanings usually associated with need-satisfaction or positive value, since such scales as *good-bad*, *pleasant-unpleasant*, etc., had high loadings. For the purposes of the present experiment the three adjectives *near* (loaded .41 on the evaluative factor), *high* (.59), and *bright* (.69), were selected as signs of goodness, pleasantness, and need-satisfaction, and their opposites, *far*, *low*, and *dark*, were selected as signs of unpleasantness, etc.

An experimental manipulation corresponding to each of these pairs of adjectives was also selected—an object moving toward and away from *S* for *near-far*, arrows pointing up and down for *high-low*, and a light going on and off for *bright-dark*. The adjectives and their corresponding manipulations are presented in Table 1 together with the predictions about guessing behavior.

In summary, the relationship between the adjectives and positive value is based upon a factor analysis of meaning by Osgood and Suci (10). The assumption that the experimental manipulations reflect the meanings of the corresponding concepts is partially inferred from previous experiments relating such manipulations to conceptual meaning.

¹ This research was completed while the authors were at The Menninger Foundation as USPHS Postdoctoral Clinical Research Fellows of the National Institute of Mental Health. The study was partially supported by a research grant (M-715C) from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service. The authors wish to thank Drs. Gardner Murphy, George Stern, and Riley Gardner for their suggestions in the preparation of the manuscript.

TABLE 1
OUTLINE OF THEORETICAL MODEL LEADING TO PREDICTIONS IN FOUR EXPERIMENTS

Experiment	Experimental Manipulations	Semantic (Adjective) Code	Semantic Associations with Code (Osgood & Suci)	Perceptual Expectancy	Predicted Guessing Behavior
I, II	Object Moving Toward S Object Moving Away from S	Near Far	Good, Pleasant, Valuable Bad, Unpleasant, Worthless	Many Few	Overguess Underguess
III	Arrow pointing up Arrow pointing down	High Low	Good, Pleasant, Valuable Bad, Unpleasant, Worthless	Many Few	Overguess Underguess
IV	Light Turned On Light Turned Off	Bright Dark	Good, Pleasant, Valuable Bad, Unpleasant, Worthless	Many Few	Overguess Underguess

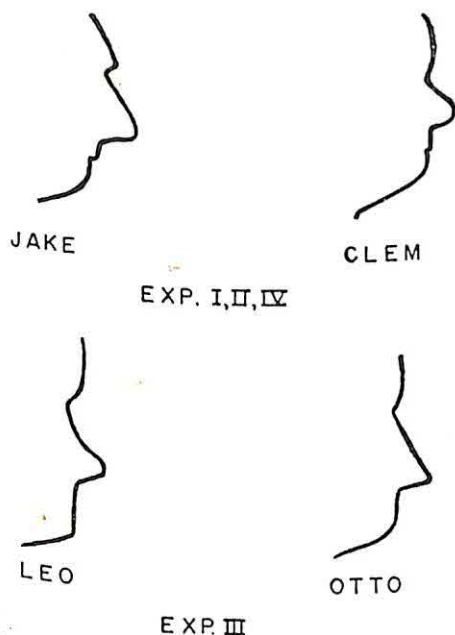


FIG. 1. PROFILES USED IN FOUR EXPERIMENTS

(12). The experimental manipulation is thought to represent a meaning similar to that of the corresponding adjective, e.g., "movement toward" signifies "near," and the manipulation is thus related to need satisfaction and positive value via the concept's semantic relation to similar pleasant and positive meanings.

METHOD

Subjects. In Exp. I, the 17 Ss were housewives, businessmen, secretaries, and psychologists. Nine Ss were assigned to a condition in which objects were moved toward S when one profile appeared and away from S when another profile appeared. The remaining eight Ss were assigned to a counterbalanced condition in which the movements toward and away corresponded to the opposite profiles. The objects were never given to S but were merely moved toward him. In Exp. II the Ss were 25 female secretarial students, 18 to 21 years of age, from a local business school.

Five were assigned to each of four experimental conditions and five to a control group. In Exp. III the Ss were again 22 female secretarial students, 8 of whom were assigned to one experimental condition and 8 to the other; a control group of 6 Ss received no experimental treatment. In Exp. IV the Ss were also 18 female secretarial students, nine of whom were again assigned to each of two counterbalanced conditions. The Ss in each experiment were randomly assigned to conditions.

Materials. In Exp. I thirty 3 x 5 index cards were used, on each of which appeared a schematic drawing of a right-pointing face. There were two such profiles; one was called "Clem" and the other "Jake" (see Fig. 1). Each profile appeared on 15 of the cards, arranged in a different random order for each S. In the experimental procedure one profile was associated with an object moving toward S and the other with an object moving away from S. The objects used in these movements were quarters, nickles, pennies, salt shakers, paper clips, hairpins, buttons, and burnt matches.

Experiment II involved the same cards used in Exp. I. Only one coin, a half-dollar, and one nonmoney object, a burnt match, were used in the movements toward and away from S.

In Exp. III two decks of 30 index cards with similar profile drawings were used. On 15 cards in each deck a profile called "Otto" was drawn and on the other 15 a profile called "Leo" was drawn (See Fig. 1). Again the two profiles were randomly mixed for each S. In one set of cards, Deck A, an arrow pointing up was drawn on cards bearing the Leo profile and an arrow pointing down on cards bearing the Otto profile. In the second set, Deck B, the direction of the arrow was reversed for Otto and Leo.

Experiment IV involved the same decks used in Exps. I and II. A small wooden box containing a seven-watt light was used as an association stimulus, the light being turned on and off with a foot pedal. The light was turned on for one profile and off for the other, and the association of a profile with "light on" and "light off" was reversed in a counterbalanced condition. Each experiment consisted of a training series and a guessing series. In the training series S was told to watch while E turned the cards over one at a time, exposing a profile and saying its name aloud. Then in Exp. I and II E moved an object toward S during the exposure of one profile and away from S during the exposure of the other. In Exp. III E merely

turned the cards over, named the profile aloud, and went on to the next card. In Exp. IV *E* turned the light on when one face was exposed and off when the other face appeared.

In the guessing series of each experiment *E* shuffled the deck and then instructed *S* to guess which of the two faces would appear on the next card. When *S* had made his guess, *E* would turn the card over exposing the profile, and *S* would make a guess about the next card. The same deck was used for both the training and guessing phases of each experiment except Exp. III, in which case a deck containing the same profiles but without arrows was used for guessing.

RESULTS

Experiment I. An analysis of variance was done on the number of guesses each *S* made for the profile associated with "toward" movements. This 2 x 2 analysis tested whether the movement of "money" objects had a greater effect than the movement of "non-money" objects and whether there were any differences between profiles. The results indicated no significant differences between money and nonmoney, between the effects of the two profiles, or for the interaction of the two variables.

Since there were no significant differences among these experimental variables, results for the different groups of *Ss* were pooled to test for differences between guesses about toward- and away-associated profiles. The two means and the *t*-test between them are shown in Table 2. This table indicates that *Ss* overguessed the occurrence of toward-associated profiles at the .01 level of confidence. Fourteen of the 17 *Ss* tested were in the predicted direction with only three in the opposite direction.

Experiment II. The analysis of this study parallels that of Exp. I. Sixteen *Ss* overguessed

TABLE 2

FREQUENCY OF GUESSES FOR PROFILE ASSOCIATIONS IN RELATION TO CONTRASTING MANIPULATIONS

Experimental Manipulation	Mean	σD	<i>t</i>	<i>p</i>
Expt. I (<i>N</i> = 17)				
Toward	16.88	.968	3.88	< .01
Away	13.12			
Expt. II (<i>N</i> = 20)				
Toward	16.00	.788	2.54	< .05
Away	14.00			
Expt. III (<i>N</i> = 16)				
Arrow-up	16.31	.871	3.01	< .01
Arrow-down	13.69			
Expt. IV (<i>N</i> = 18)				
Bright	16.33	1.073	2.48	< .05
Dark	13.67			

the profile associated with toward-movements, 3 *Ss* underguessed, and there was 1 tie. Again, an analysis of variance on the number of guesses for the toward-associated profiles showed no significant difference between movement of money and nonmoney objects, between profiles, or for the interaction of money-nonmoney and profiles. Again results were pooled over these experimental conditions to test differences between toward- and away-profiles. The means for these two variables and the *t* test between them are shown in Table 2. The toward-associated profile was significantly overguessed at the .05 level of confidence. There was no significant difference between the number of guesses for the two profiles in the control group.

Experiment III. Since no significant differences were found in a *t* test between "Leo" and "Otto" profiles, data for both groups were pooled to test the significance of the difference between "arrow-up" and "arrow-down" associations. The *t* test between these two means (Table 2) is significant at the .01 level in the direction of overguessing the "arrow-up" associated profiles. Twelve *Ss* overguessed the frequency of occurrence of the profile associated with the arrow up, 2 *Ss* underguessed it, and there were 2 ties. There was no significant difference between the number of guesses for the two profiles in the control group.

Experiment IV. Again no significant differences were found between the "Clem" and "Jake" profiles, and the data for the two counterbalanced profile groups were pooled to test the effect of "light-on" and "light-off" associations. The means for these two associations are shown in Table 2 with the corresponding *t* test. Again it was found that *Ss* significantly overguessed the occurrence of profiles associated with the light-on condition. Ten *Ss* overguessed the profile associated with "light-on," 4 *Ss* underguessed it, and there were 4 ties.

DISCUSSION

A discussion of the results and their implications can take several directions. Let us consider the results first in terms of the possible operation of autistic factors and then in terms of their implications for Osgood's mediation theory of meaning (9).

The results indicate that *Ss* made more

guesses for profiles associated with movements toward them, arrows pointing up, and lights turning on. These results are interpreted as indicating that the perceptual expectancy for the occurrence of profiles has shifted from equal expectancy, as shown in the control group, to a preponderance for the positively-associated profiles in the experimental groups. If one considers perceptual expectancy to be related to cognitive processes, then this interpretation is similar to the definition of autism given by Levine, Chein, and Murphy (7): "Autism is the movement of cognitive processes in the direction of need-satisfaction."

The movement of cognitive processes in the present experimental situation refers to the shift in perceptual expectancies as reflected in the preponderance of responses for positively-associated profiles. This "movement" differs from the type considered in previous autism studies (6, 7, 11), in which attempts were made to elicit *perceptual* shifts in the direction of rewards and away from punishments. The above conception of autism does not completely describe the results of the present experiments, since it is difficult to relate "arrows pointing up" and "lights going on" to need-satisfactions directly. In order to describe the results adequately in terms of an autism theory, a broader conceptualization is required. Some qualification of the phrase "in the direction of need-satisfaction" seems called for. One possible reformulation might define autism as the movement of cognitive processes in ambiguous situations in the direction of meanings or interpretations which have positive value for the organism. Of course, operational specification of the conditions under which cognitive processes "move" and what is meant by "movement" in individual experiments should also be made. In the present studies the movement is identified with a shift in response frequencies in favor of positive associations.

An alternative and not necessarily conflicting interpretation of the present results can also be made in terms of differential emphasis; i.e., movements toward S, arrows pointing up, and lights going on might be considered to have a greater emphasizing effect than the opposite manipulations. However, the use of an explanation in terms of emphasis requires some basis for differentiating the relative

emphatic value of arrows pointing up v.s. arrows pointing down, etc. The hypothesis nevertheless seems amenable to experimental appraisal.

If it is correct to conceive of the association stimuli used in the present experiments as translations of semantic codes into experimental manipulations, then the results also support Osgood's mediation theory and are consistent with his factor analysis of meaning (10). The evidence presented here suggests a behavioral reality for the semantic relationships obtained from Osgood's judgmental studies. Thus, when a scale has a significant loading on the evaluative factor, for example, the verbal and perceptual signs associated with it appear to function evaluatively in behavior.

SUMMARY

As a test of predictions derived from Osgood's (9) theory of meaning and Murphy's (8) theory of autism, four experiments in guessing behavior were conducted. Adjectives highly correlated with goodness and positive value (10) were selected, and translations of these semantic codes into experimental manipulations were specified. Movements of objects toward and away from S were chosen to correspond to *near-far*, arrows pointing up and down to *high-low*, and a light going on and off to *bright-dark*. A random series of two profiles was presented to S, each profile being associated with one of the experimental manipulations. Later a different random series of the same two profiles was presented, and S was asked to guess the order in which they would appear. Profiles associated with operations reflecting the adjectives *near*, *high*, and *bright* were significantly overguessed while their opposites were underguessed.

REFERENCES

1. ALLPORT, F. H. *Theories of perception and the concept of structure*. New York: Wiley, 1955.
2. ANSBACHER, H. Perception of number as affected by the monetary value of the objects. *Arch. Psychol.*, 1937, No. 215.
3. BRUNER, J. S., & GOODMAN, CECILE G. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33-44.
4. BRUNER, J. S., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203-208.

5. GILCHRIST, J. C. & NESBERG, L. S. Need and perceptual change in need-related objects. *J. exp. Psychol.*, 1952, **44**, 369-376.
6. JACKSON, D. N. A further examination of the role of autism in a visual figure-ground relationship. *J. Psychol.*, 1954, **38**, 339-357.
7. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283-293.
8. MURPHY, G. *Personality: A biosocial approach to origins and structures*. New York: Harper, 1947.
9. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197-237.
10. OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325-338.
11. SCHAFER, R., & MURPHY, G. The role of autism in a visual figure-ground relationship. *J. exp. Psychol.*, 1943, **32**, 335-343.
12. SOLLEY, C. M. Problem-solving difficulty as function of deviation of "meaning" of physical cues from expected "meaning." *J. gen. Psychol.*, in press.
13. SOLLEY, C. M., & LEE, R. Perceived size: Closure versus symbolic value. *Amer. J. Psychol.*, 1955, **68**, 142-144.

Received February 20, 1956

ACCURACY OF INTERPERSONAL PREDICTION AS A FUNCTION OF JUDGE AND OBJECT CHARACTERISTICS¹

BELA O. BAKER AND JACK BLOCK

Institute of Personality Assessment and Research, University of California, Berkeley

THE processes involved in identifying the reactions and characteristics of other people have been of persistent interest to psychologists for both theoretical and practical reasons. Many recent studies in this area have employed the technique of interpersonal prediction to secure an objective measure of accuracy of understanding. Largely neglected in such studies has been the influence of the person whose behavior is predicted, although it is widely acknowledged that the nature of the social object is an important determiner of accuracy.

The chief aim of the present study was to see whether the level of accuracy achieved in an interpersonal prediction task is predictably affected by systematically selecting the objects of prediction. To this end the objects used were selected to represent three levels on each of two dimensions expected to be relevant to accuracy of prediction.

The first independent variable was *favorableness of self-description*, by which is meant the degree to which an individual portrays himself

¹ This research was supported in part by the United States Air Force under Contract No. AF 18 (600)-8, monitored by Technical Director, Detachment #7 (Officer Education Research Laboratory), Air Force Personnel and Training Research Center, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

Personal views or opinions expressed or implied in this publication are not to be construed as necessarily carrying the official sanction of the Department of the Air Force or of the Air Research and Development Command.

The study was carried on at the Institute of Personality Assessment and Research, University of California, and was also supported in part by the grant of the Rockefeller Foundation to that Institute. Since assessment is a group endeavor, it scarcely needs to be indicated that the authors are greatly indebted to the staff of the Institute who conducted the two assessments from which the data are here drawn and who also contributed many helpful suggestions concerning the design and presentation of this research. We are grateful especially to Drs. H. G. Gough, R. B. Holt, and F. Vanasek who made detailed criticism of an early draft of this manuscript. Many of their suggestions have been incorporated in this final version.

in a socially desirable light. This variable has been noted previously by Spanner (16) and by Stone and Leavitt (17) as influencing accuracy scores. Spanner found that the most self-critical of four volunteer social objects was less accurately predicted than the other three. Stone and Leavitt employed what they term an "intermediary key" to partial self-favorableness out of the correlation between predicted and actual responses to a personality questionnaire. They found that both general level of accuracy and generality in the sense of correlations between scores achieved for different objects were decreased. Their explanation is that a judge who tends to attribute favorable self concepts to other people will score a number of adventitious hits when predicting the responses of an object who describes himself favorably, regardless of the degree of actual understanding he has of the object. These findings clearly suggest the hypothesis that social objects who are most self-favorable will be most predictable while social objects who are least self-favorable will be least predictable.

The second independent variable studied was *ego control*. Ego control refers to the individual's characteristic means of modulating or coping with his internal needs or drives. Subjects characterized by three kinds of ego-control systems were represented in the present study: Over-controlled (OC), Appropriately-controlled (AC), and Under-controlled (UC). Descriptions of the behavioral properties of these personality systems are given elsewhere (1, 4, 5, 6). Summarizing briefly, OCs bind their tensions excessively and are relatively constricted, categorical, and repressing individuals; UCs tend toward direct and unmodulated expression of their impulses and are relatively expansive, distractable, loosely organized individuals; and ACs selectively bind and gratify their needs in accord with the social reality in which they function.

If it is assumed that the greater the consistency or channelization of an individual's behavior, the more completely he is predictable,

then the rank order of accuracy scores for these three groups of objects can be predicted. OCs, who are conceptualized as developing relatively set ways of behaving, should be most predictable; UCs, who are characterized as manifesting frequent inconsistencies, should be least predictable; while ACs should be predictable at an intermediate level.

A further hypothesis was made concerning the accuracy of judges classified in terms of ego control. Our expectation was that OC and UC judges would be less accurate than AC judges. Our assumption in making this prediction was that two components are required for accurate social prediction, sensitivity of social perception, and the ability to integrate perceptions into a formulated prediction scheme. OCs are considered to be relatively excluding of the potential social field while UCs, because of their inability to establish structure or a frame of reference for themselves (4), are relatively unable to construct a coherent basis for prediction. Since ACs possess both required characteristics, they should be superior to both OCs and UCs.

METHOD

Stimulus material available to judges. In the Spring of 1953, 100 Air Force captains were studied during a three-day assessment program at the Institute of Personality Assessment and Research. A standardized stress interview, conducted by a grim and unresponsive examiner, was the initial procedure experienced by each officer. The interview was conducted under floodlights before a group of observers with a sound movie camera operating² in plain view of the subject. The schedule of questions started with, "How does it feel to be in the movies?" and moved on to such questions as, "What sort of a person are you?", "What do you consider to be your main personality assets?", and "What do you consider to be your main personality handicaps?" The questions elicited rather rich samples of the expressive and coping behavior of the subject.

Nine officers were selected from the pool of 100 potential social objects on the basis of the considerations indicated below and the films of the interviews served as the standard stimulus material in the study. The films were approximately two and one-half minutes in duration and were of good quality both visually and with respect to sound reproduction.

Nature of the required prediction. In the course of the assessment, each officer Q-sorted a set of 54 adjectives along a 9-point scale so as to describe himself. The task set for the judges was to predict these self-descriptions. To make the judges' task somewhat lighter, several categories were combined, reducing the sorts to a

5-point scale with the fixed frequency distribution: 2, 10, 30, 10, 2. The accuracy score employed was the correlation coefficient between the judge's predicted Q sort for an object and that object's self-sort.

The self Q sort was chosen for prediction because the content of the movies appeared to be highly relevant to the task of evaluating the object's self-percepts and because the prediction of some form of self-description is a widely used experimental task. An important methodological advantage is that the response set variations in elevation and dispersion which Cronbach (9) has noted as confounding the discrepancy score are eliminated by the use of a forced distribution.

The Objects

Measurement of object self-favorableness. To assess self-favorableness, six psychologists were requested to sort the 54 adjectives to predict the self-description of an individual characterizing himself in the most favorable light. The average intercorrelation among the six maximally self-favorable sorts was .73, indicating that a composite of these six Q sorts has an estimated reliability (Spearman-Brown correction) of .94.

This maximally self-favorable Q sort was employed as a reference key with which the self-sorts of the 100 officers were correlated. A high correlation implied an officer was describing himself favorably; a low correlation that he was ascribing socially undesirable characteristics to himself. Selection of social objects in terms of self-favorableness was based upon this set of correlations treated as scores.

Measurement of ego control among the objects. At the end of the three-day assessment program, eight staff psychologists codified their personality evaluations of each of the officers by means of a 76-item Q sort. The eight separate Q sorts for each subject were combined into one composite Q sort which thus represented the staff consensus as to the nature of the subject's personality. The sample of officers was divided into equivalent groups of 50 and, within each group, the consensus Q descriptions were intercorrelated. These correlations between persons were then cluster-analyzed (18).

A number of personality "types" was found, five of which were observed in both 50 × 50 analyses. The cross validated types were combined. When evaluated in terms of test and personality scale measures, the cluster-types isolated by the inverse analyses proved to be reliably different among themselves, and in a fashion compatible with the defining characteristics of their

TABLE 1
CORRELATIONS BETWEEN SELF Q SORTS AND THE FAVORABLENESS KEY FOR THE NINE SOCIAL OBJECTS

Self-favorableness	Ego Control		
	Over-Controlling Objects	Appropriate Controlling Objects	Under-Controlling Objects
High Self-favorableness			.64
Middle Self-favorableness	.69	.61	.36
Low Self-favorableness	.42	.43	.03
	.08	.08	

² Dr. Victor B. Cline served as cameraman during these interviews.

modal Q sorts. A fuller explanation of the procedure and results of these analyses is presented elsewhere (2).

On the basis of the Q-sort descriptions and other measures characterizing them, three of the groups were designated as fitting the categories of over-, under-, and appropriate-control (*N*s of 26, 26, and 21, respectively). These groups constituted the populations from which the social objects were drawn.

Specification of the social objects. Representatives of the three ego-control groups were selected at each of three levels of self-favorableness, making a total of nine social objects. Table 1 shows the degree of matching for favorableness within the three ego-control groups. The ages of the social objects varied from 29 to 44, with a mean of 34.4. Their socioeconomic origins are best estimated as upper-lower or lower-middle class.

The Judges

The sample of judges. During the winter of 1954-55, 70 applicants to medical school underwent an 18-hour assessment at the Institute. These assesseees were preponderantly undergraduate, although a number were graduate students or in the nonacademic community. They ranged in age from 19 to 33, with a mean at 22. Thus, the judges were substantially younger than the social objects. Socioeconomically, the group was essentially of lower-middle to middle-middle class origin, somewhat higher than the social objects.

Procedure with respect to judges. The procedure was administered to groups of five judges at a time in two 40-minute sessions. Before viewing any films, the subjects Q-sorted the 54 adjectives so as to describe themselves. They were then instructed and four films were shown, with suitable intervals for the judges to indicate their predictions. After an interval of two hours, the remaining five films were shown. The films were presented in a standard order such that no social object followed immediately after another object from the same ego-control or self-favorableness group.

Measurement of ego control among judges. At the end of the assessment, seven staff psychologists codified their personality evaluations of each subject-judge by means of a 115-item Q set. The seven separate Q sorts describing each subject-judge were combined into one composite Q sort representing the staff consensus as to the subject's personality.

Two of the Q items that were evaluated by the assessors for each subject were, "Tends toward over-control of his needs and impulses; binds his tensions excessively; delays gratification unnecessarily," and "His various needs tend toward relatively direct and uncontrolled expression; unable to delay gratification; acts out." The relative weighting of these two items provided the basis for categorizing judges into one of the three ego-control groupings. For each subject-judge, his Q score on the under-control item was subtracted from his Q score on the over-control item. A distribution of difference scores was thus obtained where high positive scores implied over-control and negative scores under-control.

Four judges were eliminated from the sample for various reasons: one because he grew up in a foreign country, one because he was extremely deviant from the rest of the sample on a variety of assessment measures, and two because they appeared to belong more

clearly to an ego-control category not here used—that of bimodal control. The remaining 66 judges were divided into three groups on the basis of the distribution of difference scores. The 22 individuals with the highest positive difference scores were designated as the group of over-controlling judges; the 22 individuals with the highest negative difference scores constituted the group of under-controlling judges; and the 22 remaining subjects in the midrange were classified as appropriately-controlling judges.

Although the judges were categorized on the basis of only two Q items, other items which covaried with these two could be examined to give a more comprehensive picture of the differential characterizations of the groups. It was not possible to make a direct comparison of the Q-sort descriptions of the judges and objects, since different Q decks were used. Inspection of the items differentiating among the three judging groups yields a subjective impression that the dimension involved is quite comparable to that distinguishing among the object groups.³ The OCs are in both cases described as relatively constricted, passive, conforming individuals; the UCs as more ascendant, expressive, and rebellious; and the ACs as more moderate individuals who share the desirable but not the undesirable traits of the extreme groups.

RESULTS

Accuracy was scored by correlating predicted Q sorts with the appropriate object self-sorts. The accuracy scores were then treated by an analysis of variance design⁴ adapted from

³ Tables of the items significantly differentiating among UCs, ACs, and OCs in the samples of objects and judges have been deposited with the American Documentation Institute. Order Document No. 5035 from ADI Auxiliary Publications Project, Photoduplication Service, Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to chief, Photoduplication Service, Library of Congress.

⁴ It should be noted that there is a problem involved in applying analysis of variance to these data. For practical reasons it was necessary to present the stimulus materials to the subjects in a fixed, rather than random, order. Consequently, the possibility that differences between object groups are confounded by order effects must be considered. The order used placed one member of each object classification in each third of the presentation series. Thus the ordinal positions of objects classified by ego-control were: OCs—1, 6, 8; ACs—3, 4, 7; UCs—2, 5, 9. For objects classified by self-favorableness the ordinal positions were: High—1, 4, 9; Middle—3, 5, 6; Low—2, 7, 8. Our trust was that this approximate balancing of serial position within groups would largely exclude order effects from the main comparisons and the simple interactions.

Since the correlations which comprised the raw accuracy scores appeared by inspection to form a reasonably symmetrical distribution, they were coded for this analysis by simply dropping the decimal point and adding a constant of 30 to remove negative numbers.

TABLE 2
ANALYSIS OF VARIANCE OF PREDICTIVE
ACCURACY SCORES

Source	df	Sum of Squares	Error Term	F
Between personality-style of judges (JP)	2	2,596.46	b	3.05*
Error b	63	26,843.58		
Between subjects	65	29,440.04		
Between personality-style of objects (OP)	2	4,909.12	w ₁	6.07†
Between self-favorableness of objects (OF)	2	36,568.50	w ₂	94.48†
Interactions:				
OP × OF (or OID-Residual individual differences among objects)	4	151.08	w ₃	—
OP × JP	4	253.94	w ₁	—
OF × JP	4	958.04	w ₂	1.24
JP × OID	8	687.53	w ₃	1.68
Error w	504	88,205.56		
Error w ₁	126	50,915.38		
Error w ₂	126	24,383.25		
Error w ₃	252	12,906.93		
Within subjects	528	131,733.77		
Total	593	161,173.81		

* For 2 and 60 degrees of freedom, $F_{.10} = 2.39$, $F_{.05} = 3.15$.

† For 2 and 120 degrees of freedom $F_{.005} = 5.54$, $F_{.001} = 7.31$.

one described by Lindquist (13, pp. 292-297). Table 2 summarizes the results. Two of the main effects were highly significant (p less than .005), while the other was of borderline significance (p approximately .06). None of the interactions was significant.

The social objects expressing the most favorable self-descriptions were most readily predicted, the average accuracy score achieved for them being .42. The objects at the middle and lowest levels of self-favorableness did not differ reliably in predictability. Indeed, the low favorableness objects were somewhat more predictable than those in the middle range, average accuracy scores for the groups being, respectively, .24 and .27. Our hypothesis concerning the effect of self-favorableness on predictability is thus partially supported and partially infirmed.

The over-controlling social objects were most accurately judged, the average accuracy score achieved for them being .35. The appropriate- and under-controlling objects, for whom average accuracy scores were .28 and .29 respectively, did not differ in predictability. Our

hypothesis concerning this effect is thus also partially supported and partially infirmed.

The AC judges performed at the highest level of accuracy, achieving an average score of .33, and the OC judges were least effective, achieving an average score of .28. The intermediate UC judges, with average scores of .31, were not significantly different from the judging groups at either extreme. Our hypothesis concerning the judges' relative accuracy thus received some support.⁵

DISCUSSION

Accuracy of prediction as a function of the judge. Many current theories of personality suggest that the ability to understand others is an important feature of good adjustment. From this point of view, one would expect that AC judges, who are conceptualized as responding more adaptively to reality, would be more accurate than either OCs or UCs in their accuracy of prediction. And in fact, this was the case, although the magnitude of the differences is small. Since this finding is in general agreement with previous results (e.g., 8, 10, 12), it will not be discussed further except to point out that the result does not necessarily imply AC judges, or better adjusted judges, are more "understanding" of others. At this stage of knowledge, the result simply indicates that these judges are more accurate predictors of others, regardless of how that accuracy is achieved.

The question can be raised at this point as to whether the various groups of judges differed in the manner in which they made predictions. In exploring this question a number of measures were defined, two of which are mentioned here because of their particular relevance. The first measure, labeled "ascription of self-favorableness," is an index of the degree to which favorable self-opinions are attributed to others, a measure which was defined as the correlation between the composite of an individual's predictions and the favorableness key

⁵ It should be noted here that the observed level of significance for this hypothesis is an underestimate. The over-all F -test, in the situation where a prediction is supported, provided too conservative an estimate of the true level of significance. An allowance for the ordering of the three groups as predicted is difficult to decide upon. A halving of the obtained p level (.06) is still somewhat conservative but is closer to the correct significance value.

mentioned earlier. An analysis of variance of these scores indicated that there were no significant differences among the groups of judges.

The second measure, tentatively called "self-generalization," was defined as the correlation between the composite of a judge's predictions and his own self-description. An analysis of variance of these scores again indicated that the differences among the groups of judges were not significant. The variations in accuracy described above, therefore, cannot be explained either in terms of variations in the tendency to ascribe favorable self-opinions to others or in terms of variations in the tendency to ascribe one's own self-perception to others.

Accuracy of prediction as a function of the object. The degree of favorableness with which the social objects described themselves was found to be strongly related to the accuracy with which the judges were able to guess their self-descriptions. However, our hypothesis concerning this effect was not fully supported, for, although the highly self-favorable objects were best predicted, the objects with middle and low degrees of self-favorableness did not differ in their predictability and in fact were ordered differently than had been anticipated.

One partial explanation of this result is to be found in the fact that all the judges tended to predict relatively favorable self-descriptions, the average correlation between the composites of individual judge's predictions and the favorableness key being .39, with a standard deviation of .15. Consequently, the degree to which an object's self-description correlated with the favorableness key helped determine the accuracy with which it was predicted. Such a finding—which is not new with us (cf. 16, 17)—is, upon reflection, to be expected from the context of prediction experiments. There is a good deal of social pressure toward the suppression of unfavorable views of one's fellow man. We may cite here the homily, "If you can't say something good about a person, don't say anything." Some of our subject-judges may have been more than ordinarily aware of this pressure because of the fact that they were undergoing an assessment and may have felt that it would reflect ill upon them if they ascribed unfavorable characteristics to the objects. Others may have reasoned (and not too inaccurately) that the objects themselves, since they were also assesseees, would have been

concerned about making good impressions at the time they sorted the adjectives. In general, the fact that the experimental situation was ambiguous and even a little threatening from the subjects' point of view is likely to have promoted cautious—i.e., favorable—responses.

Beyond the common tendency toward caution evoked by the role of being a judge, there probably are reliable individual differences in the extent to which judges adopt the approach of ascribing favorableness. These individual differences among judges may well be in themselves fruitful objects of study. But the degree to which they should be allowed to enter into measures of judging ability must be determined by the interests of the investigator. If the research focus is on accuracy of prediction per se, then all mediators of accuracy must be accepted as relevant. If, however, the orientation is not toward predictive accuracy per se, but rather toward a notion of "interpersonal acuity," "social sensitivity," or "understanding," then the judges' tendencies to make favorable predictions may be interpreted as confounding, and this source of variance should in some way then be controlled or partialled out.

De-emphasizing the effects of object self-favorableness may be accomplished in a number of ways. Some form of partial correlation may be employed to eliminate the ascription-of-favorableness effect, as in the Stone and Leavitt (17) study, leaving a reduced score but one free of this particular component. But as perhaps the better technique, judges may be prevented from predicting favorableness by the stratagem of employing a forced choice procedure where the alternatives are matched for social desirability, as in, for example, the Edwards Personal Preference Schedule (11).

The influence of the objects' ego-control style on predictability is not to be so neatly accounted for as is the influence of self-favorableness. Because our hypothesis concerning this effect was only partially supported, the result probably does not warrant extensive discussion. While the greater predictability of OC objects supports the conceptualization of over-control previously advanced, the meaning of the equivalence, even slight reversal, of the predictability of AC and UC objects remains obscure.

Although the finding itself does not clarify the process of understanding, it nevertheless

demonstrates that predictive accuracy scores can be manipulated by varying an inferred central dimension. This demonstration takes on more force when one considers the really startling (at least to us) finding that the residual individual differences among objects—i.e., all individual differences among objects over and above those accounted for by the main effects—failed to elicit any significant variance in accuracy of prediction. Thus by systematically varying social objects in terms of self-favorableness—a dimension defined by explicit features of the objects' self-descriptive responses—and ego-control—a postulated structural characteristic of personality—all the variance in object predictability in this experimental situation has been exhausted.

While this finding impresses us as noteworthy, we would hesitate to make too much of it until it is repeated. Self-favorableness is an obvious and easily controlled characteristic of verbal descriptions, the relevance of which is greatly amplified within the context of social prediction experiments. Ego control is a personality dimension underlying a broad variety of behaviors. That the two together should account for the bulk of the variance in predictability is not implausible. Even if this is the case, however, it still does not exclude the possibility that other dimensions overlapping with the ones used here may bear a significant relationship to predictability.

Favorableness, as indicated earlier, is a conveniently manipulated characteristic of verbal descriptions, whether of self or of others. Such is not the case with ego control as a determinant of predictions. It might be possible to derive a key from the present data to reflect those differences among objects attributable to their different ego-control patterns. This key could then be used to partial out of the accuracy scores those differences among them attributable to the objects' ego-control styles. But clearly, such an after-the-fact key would have little general application, since it could not be defined independently of the objects. Experimental control over the effects of this variable cannot be attained, therefore, by the expedient of using partial correlations.

This fact, of course, poses no problems for studies of predictability and merely complicates the study of the mediation of accuracy by the judges' response tendencies. For studies of

judging ability, however, the indication seems to be that careful attention to the problem of sampling is essential if results are to be generalized beyond the specific experimental situation. In this respect, Brunswik's thesis (7) applies—the same considerations that govern the sampling of judges should apply to the sampling of objects.

Dimensional extremeness as a possible determinant of predictability. In the present experiment, the social objects in the midrange of the independent variables studied, i.e., the middle level of self-favorableness and the AC level of ego control, were least accurately predicted. In each case the difference in predictability between these middle groups and the extreme group which fell next in order of predictability was slight and nonsignificant. But the ranking itself makes implausible the assumption that there is an ordinal relation between predictability and either of the experimental dimensions. We would expect that the inclusion of more extreme representatives of low self-favorableness and under-control among the social objects would make the curvilinear trend more pronounced.

These findings, in conjunction with our knowledge of the process of concept formation and the development of frames of reference (14, 15), suggest that judges seek clear and unambiguous categories ("stereotypes") as aids in prediction. Socially valid personality categories, informal though they may be, provide an economical and integrative framework from which specific predictions may be made. "Likes himself"—"doesn't like himself" or "rigid"—"impulsive" are readily available organizing principles with a good deal of social utility.

The social validity of these categorizing principles, however, stems from the behavioral consistency of individuals at extreme positions on the continua used. In predicting the explicit self-descriptions of self-ambivalent objects or in assessing the conditional behavior of appropriate-controllers, simple notions of consistency no longer apply. Consequently, accuracy of prediction suffers.

SUMMARY

A sample of nine social objects was selected so as to be systematically varied on two dimensions: ego control and self-favorableness. Sixty-

six subjects viewed sound movies of the behavior of each object in a stress interview and attempted to predict his self Q sort of 54 adjectives. The judges were also classified on the dimension of ego control. An analysis of variance of the accuracy scores indicated that significant variance was associated with ego-control style of the judges, and with both ego-control style and self-favorableness of the social objects. Residual individual differences among the objects were *not* significant. The results are discussed in terms of their implications for the methodology of social perception studies, with emphasis on the importance of a carefully considered selection of standard social objects.

REFERENCES

1. BLOCK, J. An experimental investigation of the construct of ego-control. Unpublished doctor's dissertation, Stanford Univer., 1950.
2. BLOCK, J. A differentiated approach to the officer selection problem. *Officer Educ. Res. Lab., USAF Pers. Train. Res. Cent. Res. Rep.*, March 1955, Project No. 7730, Contract No. AF18(600)-8 (Rep. No. 8). Pp. 1-34 (dittoed).
3. BLOCK, J. & BLOCK, JEANNE. An investigation of the relationship between intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1951, **19**, 303-311.
4. BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254-259.
5. BLOCK, JEANNE. An experimental study of a topological representation of ego-structure.

- Unpublished doctor's dissertation, Stanford Univer., 1951.
6. BLOCK, JEANNE, & MARTIN, B. Predicting the behavior of children under frustration. *J. abnorm. soc. Psychol.*, 1955, **51**, 281-285.
7. BRUNSWIK, E. *Systematic and representative design of psychological experiments*. Berkeley: Univer. of California Press, 1947.
8. CLINE, V. B. Ability to judge personality assessed with a stress interview and sound-film technique. *J. abnorm. soc. Psychol.*, 1955, **50**, 183-187.
9. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177-193.
10. DYMOND, ROSALIND F. Personality and empathy. *J. consult. Psychol.*, 1950, **14**, 343-350.
11. EDWARDS, A. L. *The Edwards Personal Preference Schedule*. New York: Psychological Corp., 1954.
12. GAGE, N. L. Accuracy of social perception and effectiveness in interpersonal relationships. *J. Pers.*, 1953, **22**, 128-141.
13. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
14. NEWCOMB, T. M. *Social psychology*. New York: Dryden, 1950.
15. SHERIF, M. *The psychology of social norms*. New York: Harper, 1936.
16. SPANNER, M. Similarity, identification, and distortion as factors in the prediction of personality characteristics. Unpublished doctor's dissertation, Univer. of California, Berkeley, 1954.
17. STONE, G. C., & LEAVITT, G. S. Generality of accuracy in perceiving standard persons. Paper read at Midwest. Psychol. Ass., Columbus, Ohio, May, 1954.

Received February 20, 1956

DIFFERENCES BETWEEN SCHIZOPHRENIC AND BRAIN-DAMAGED GROUPS IN CONCEPTUAL ASPECTS OF OBJECT SORTING

LAURENCE S. McGAUGHRAN¹

University of Houston

AND LOUIS J. MORAN²

University of Texas

SCHIZOPHRENIC thinking has been characterized by a number of investigators (e.g., 1, 2, 7, 8, 11, 14, 18, 20, 21, 22, 23) as an impairment or a deterioration in abstract ability, resulting in a "very great concreteness" (7, p. 23). Many of these same writers (e.g., 1, 2, 7, 11, 14, 23) have emphasized that the extreme concreteness of schizophrenic thinking closely resembles that of the brain-injured.

It is the contention of the investigators that studies of conceptualization have been unduly influenced by Goldstein's dichotomous formulation of "abstract-concrete." The two "conceptual levels," abstract and concrete, were originally employed by Goldstein and Gelb to distinguish the conceptual behavior of brain-damaged patients from that of normal adults. When the term, "concrete attitude," was applied subsequently by Goldstein and other investigators to characterize also the conceptual behavior of a number of groups other than normal adults, the inevitable result was to crowd together a heterogeneous group of referents for the term, without apparent univocality.

In a previous report (16), the investigators suggested that the varied uses of the concepts, abstract and concrete, represent a fusion of at least four isolable conceptual variables: (a) amount of social agreement, or communality; (b) order (or degree) of conceptual classification; (c) relative essentiality, or saliency; and (d) the presence or absence of generic terms. In that study we attempted to demonstrate that of these four meanings the one that most clearly distinguished schizophrenic and normal conceptualization was the relative amount of social agreement, or communality, for the concepts employed; on the other hand, hy-

potheses of group differences which were based upon the Goldstein dichotomous abstract-concrete formulation were not supported in a single instance.

In the present study, the hypothesis to be tested is that the second variable—order of conceptual classification—shows the greatest difference between schizophrenic and brain-damaged Ss.

METHOD

Subjects

The major samples consisted of a group of 37 male patients diagnosed as chronic paranoid schizophrenic and a group of 34 male patients who had been diagnosed as suffering from types of disorder associated with the more general diagnostic term, brain-damaged. The specific medical diagnoses included encephalopathy or chronic brain syndromes associated with encephalitis, cerebral thrombosis or arteriosclerosis, central nervous system syphilis, epilepsy, post-traumatic encephalopathy, Pick's Disease, etc.³ A group of 19 additional brain-damaged male patients whose official diagnoses included references to psychotic reactions were used in this study only to check reliability of scoring in the object-sorting procedure.

Means and standard deviations for the control variables of age, education, and estimated intelligence level are shown in Table 1. The distribution of age and education shows close agreement; neither of the mean differences for these variables approached significance. On the other hand, the significance of group mean differences in intelligence exceeds the .001 level. Since intelligence could not be controlled experimentally, and since the results of a previous study (16) had indicated a high degree of association between intelligence and conceptual behavior, analysis of covariance was employed to provide a statistical control over the effects of intelligence upon sorting behavior.

The Sorting Task and Rationale for Scoring Conceptual Areas

The sorting task used was the Rapaport modification of the Goldstein-Gelb-Weigl Object Sorting Test. The task requires 19 different sortings of a group of 33 objects familiar to everyday experience. Administra-

¹ Employed part-time at the Houston Veterans Administration Hospital, where this study was conducted.

² Study conducted while employed at the Houston Veterans Administration Hospital.

³ We wish to express our deep appreciation to Dr. Ruth M. Hubbard and other staff members of the Waco Veterans Administration Hospital for their kind cooperation in making these records available.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF DISTRIBUTIONS FOR CONTROL VARIABLES OF AGE, EDUCATIONAL LEVEL, AND ESTIMATED INTELLIGENCE

Experimental Group	N	Age		Education		Intelligence*	
		Mean	SD	Mean	SD	Mean	SD
Brain damaged	34	40.1	10.5	9.3	3.2	95.8	14.7
Schizophrenic	37	40.3	9.6	10.7	2.6	108.9	15.1

* Based upon verbal IQ scores on the Wechsler-Bellevue Scale; for the schizophrenic sample, these scores were estimated from results obtained with the vocabulary subtest.

tion was in accordance with the procedure described by Rapaport (18).

Each sorting was scored separately as occurring in one of four "conceptual areas." The basis for sorting into the four conceptual areas has been described in some detail in previous reports (15, 16). In summary, the rationale is as follows: Two variables, order of conceptual classification and extent of social agreement, are isolated in the operations of this experiment and are termed the "open-closed" and the "public-private" variables, respectively. In previous studies these variables have been found to be unrelated to each other.

Let us consider, first, the open-closed variable. The more attributes that are "bound-together" within a conceptual classification, the more "closed" the sorting, or concept, is considered to be. Thus, a concept based upon a single attribute common to all objects within the conceptual grouping would be maximally "open" (e.g., a group of various objects conceptualized as belonging to the same class because all are composed of metal). On the other hand, a concept binding together all perceptible attributes (all of the "degrees of freedom") of the objects within a conceptual grouping would be maximally "closed" (e.g., a concept based upon the absolute likeness of two sugar cubes). It can be seen that the more open a concept is, the more abstract, or generic, or the higher the order of the principle of classification that is involved.

As for the second variable, "public-privateness," a relatively high degree of publicness of a concept implies that the principle underlying the grouping is widely shared by and freely communicated to others within a social group. For example, the boundaries for a concept of "roundness" are relatively clearly denotable; the conceptual boundaries for a concept of "goodness" would be considerably less so; thus, the exact conceptual limits intended by the latter term are not easily communicated, nor are they easily anticipated by others.

⁴ A revised set of subarea categories and supplementary instructions for scoring conceptual areas as used in this study has been deposited with the American Documentation Institute. Order Document No. 5067 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

If these two variables, with their respective poles of closed-open and public-private, are visualized as intersecting in space, the resulting quadrants may be conceived to represent the following four "conceptual areas": closed-public, open-public, open-private, and closed-private. These four conceptual areas are the basic scoring categories. Examples of subcategories and typical conceptual groupings within each of the four conceptual areas are as follows:

1. Closed-public area
 - a. Absolute identity—(2 corks; basis: "exactly alike").
 - b. Species identity—(real and toy forks; basis: "both forks").
2. Open-public area
 - a. Unitary abstraction—(all metallic objects; basis: "all made of metal").
 - b. Superordination—(several toys; basis: "all these are toys").
3. Open-private area
 - a. Loose-representational—(bicycle bell, cork; basis: "you shouldn't drive when drinking").
 - b. Over-abstraction—(red ball, red paper circle, red matchbook, red sink stopper; basis: "manufactured articles").
4. Closed-private area
 - a. Species-object mediation—(bicycle bell, lock; basis: "put bell on bicycle and lock it").
 - b. Closed radial—(matches, several smoking objects; basis: "you need matches to light each of these").

It can be seen that the determination of the appropriate quadrant, or area, for a concept requires a cross classification, in which the effects of both variables can be taken into account with one notation. By summing the individual S's scores in adjacent quadrants, two additional measures, "total public vs. private" and "total open vs. closed," can be derived. These measures provide a notation of the effect of either variable, independent of the other.

One additional measure, an "autistic index," is derived by subtracting the sum of closed-public concepts from the sum of concepts scored as open-private. While the dichotomous abstract-concrete formulation would usually lead to the indiscriminate classification of concepts within both of these areas as "concrete," a relatively high negative relationship between the two classes has been found consistently.⁵

Procedure

The object-sorting records of the schizophrenic group had been scored in a previously reported study (16). The two brain-damaged samples were drawn in the following way: All cases referred for psychological observation over a number of years for whom sorting protocols were available were drawn from the file of a large hospital facility. They numbered about 100. These test records were rescored to obtain the measures described above. The official diagnosis of each case was then obtained and the 53 cases with complete

⁵ For example, with the data of the present study, the product-moment correlation between these two measures was $-.48$ for the brain-damaged group and $-.63$ for the schizophrenics.

protocols who were subsequently found to be diagnosed as brain-damaged were selected.

It was not practicable to provide as tight a control over the possibility of systematic bias in scoring as was available in the previous study (16). Therefore, several informal checks of scoring reliability were made. First, records that turned out not to be associated with a diagnosis of brain damage were compared with those of the selected sample; for each predictive measure, the average of the brain-damaged Ss exceeded that of the unselected group. Second, of the Ss who were eventually included within the brain-damaged sample, a few (about 10) were known by the scorer to have a diagnosis of brain damage because of notations on the test records. Subsequent comparison of scores for this small group against the rest of the sample indicated no systematic differences in the obtained scores. Third, the group who were subsequently identified as "brain-damaged with psychotic reaction" were found to differ considerably from the major brain-damaged sample in their obtained scores; these differences were all in a direction consistent with the hypotheses. Finally, in previous studies it has been found that with rather careful instruction and practice on scoring procedures, interscorer agreement can be raised to about 93 per cent. This compares favorably with a score-rescore agreement of 95 per cent in the present study, using 180 responses selected "blindly," and at random, from the protocols of all groups, with a time interval of about six months.

Hypotheses

In the previous study of differences in conceptualization between schizophrenic and nonpsychiatric groups, it was successfully predicted that the schizophrenic group would show "a loss of social communication (loss of 'publicness') in the formation of concepts, without evidence of a corresponding loss in high order of conceptualization (loss of 'openness')" (16, p. 45). This hypothesis grew out of a consideration of the tentative findings of one of the authors (15). In an investigation of differences in the language behavior of normal Ss, it was found that a predominant tendency to use closed concepts in object sorting was associated with "stimulus-boundness," or "reality-fixing"; similarly, a predominant tendency to use private concepts in object-sorting was associated with "autism." Thus, the hypothesis was based upon the assumption that "the peculiar forms of schizophrenic conceptualization should be more closely associated with autism and reality loss than with reality fixing and stimulus-boundness" (16, p. 46).

In the present study, the opposite assumption is made concerning the characteristic conceptualization of brain-damaged patients, viz., that it should be more closely associated with reality-fixing and "stimulus-boundness" (7) than with autism or loss of reality testing.

The general hypothesis of differences between brain-damaged and schizophrenic groups in object-sorting behavior is that the brain-damaged group has a greater deficit in higher order concepts (loss of "openness").

The specific hypotheses stated in terms of the operations of the study, are as follows: In comparison with the schizophrenic group, the brain-damaged group:

1. Employs significantly fewer open and significantly more closed concepts.
2. Employs significantly fewer (a) open-public and (b) open-private, and significantly more (c) closed-public and (d) closed-private concepts.
3. Scores significantly lower on an autistic index consisting of open-private minus closed-public concepts.

RESULTS

The results of an analysis of covariance for the conceptual area measures are shown in Table 2. The values shown in the "adjusted *F* ratio" column reflect only the residual variances attributable to experimental group differences after the covariance between intelligence and the conceptual area measures has been withdrawn. The extent of this covariation for each measure for Ss in the two groups may be seen in the column of correlation coefficients.

All six measures involved in the hypotheses indicate results consistent with the prediction of greater deficit in higher order concepts (loss of openness of conceptualization) for the brain-damaged group. Four of these group mean differences are significant at the .01 level or better.⁶ The other two (open-public and closed-private) both show considerable mean differences but also show relatively high correlation values with intelligence. The greatest *F* ratio is shown for the closed-open variable, which is the major predictive measure for the present study. No hypothesis was formulated for the public-private variable, which served as the major predictive measure in the previous schizophrenic-nonpsychiatric study; in the present study, the mean difference between groups is relatively slight. It can be seen that the predominant private conceptual area for the brain-damaged group is closed-private, as contrasted with the predominance of the open-private area in the conceptual behavior of the schizophrenic group. Apparent differences in meaning for these two kinds of private conceptualization will be discussed more fully below.

A comparison of these results with those obtained in the previous schizophrenic-nonpsychiatric study (16) indicates that the extreme group separation in every conceptual

⁶ *F* ratios to test for homogeneity of variance were computed for all measures in Table 2 for which *p* values for *F* ratios are shown. The results of these tests indicate that the hypothesis of a common population variance is tenable in all cases except for the open-private measure, although the ratio for the closed-private measure also approaches significance.

TABLE 2

SIGNIFICANCE OF DIFFERENCES IN CONCEPTUAL AREA SCORES BETWEEN SCHIZOPHRENIC AND BRAIN-DAMAGED GROUPS AND CORRELATION VALUES BETWEEN CONCEPTUAL AREA AND INTELLIGENCE TEST SCORES FOR THESE GROUPS

Measure	Schizophrenic N = 37		Brain Damaged N = 34		Adjusted F Ratio	p	r	r _p
	Mean	SD	Mean	SD				
Public	11.00	4.70	11.27	3.67	0.82	—	.42	<.01
Closed	6.05	3.35	11.18	3.29	25.34	<.001	-.32	<.01
Closed-Public	3.78	2.62	6.77	3.46	14.71	<.001	.04	—
Closed-Private	2.27	2.11	4.41	2.93	3.78	<.06	-.46	<.01
Open-Public	7.22	3.98	4.50	2.74	2.43	—	.48	<.01
Open-Private	5.73	3.95	3.32	2.48	10.11	<.01	-.13	—
Autistic Index*	11.95	5.95	6.55	5.13	22.99	<.001	-.29	<.02

* Open-private minus closed-public, plus 10.

area measure, except that of public-private, is between the schizophrenic and brain-damaged groups. In other words, both kinds of disorder produce departures from normal conceptual behavior, but apparently in opposite directions for at least most of the measures.

DISCUSSION

Methodological Considerations

A point relating to methodology concerns the effectiveness of cutting points for the separation of schizophrenic and brain-damaged groups. While the writers were not primarily concerned in this study with the matter of individual diagnosis, they do heartily concur with Yates (24) that such information is basic to this type of study. Using the sum of "closed" concepts as the major predictor of group differences and the total group mean (between scores 8-9) as the most reliable cutting point, the number of correct predictions for the brain-damaged group was found to be 26 out of 34 (77 per cent), and for the schizophrenics, 31 out of 37 (84 per cent). On the whole, it would seem reasonable to conclude that this type of analysis will not be very serviceable in individual diagnosis until the present techniques can be altered to provide a more sensitive measure of group differences in conceptual behavior.

Theoretical Considerations

In the writers' opinion, the Goldstein approach represented real progress in this much-neglected research area. But, more recently, a number of investigators (e.g., 3, 4, 5, 6, 9, 10,

12, 13, 15, 16, 17) have indicated that, if the venture is made to "cut through" Goldstein's dichotomous formulation of abstract and concrete conceptual levels, the yield is an improvement in prediction of group differences in conceptual behavior. The present investigators have attempted to show that the use of two isolated variables, previously embedded within the congeries of meanings for abstract-concrete, may yield significant quantitative distinctions in the conceptual behavior of clinical and "normal" groups, which are not evident when the sorting task data are analyzed within the duality of an abstract-concrete formulation. Thus, in a previous study (16), it was found that the schizophrenics showed a relative loss (or lack) of publicness in their conceptual behavior but that, in terms of order of classification, they actually showed more openness of conceptualization than did the nonpsychiatric group. In the present study, on the other hand, the brain-damaged group showed an extreme lack of openness of conceptualization; and with this measure the greatest group separation was between the two deviant groups, brain-damaged and schizophrenic.

The mean difference between the brain-damaged and schizophrenic groups on the public-private measure did not approach significance. This closeness of agreement is, in one sense, an artifact produced by canceling out opposite group tendencies on the separate conceptual area scores, open-private and closed-private; that is, in their private conceptualization, the schizophrenics are predominantly open and the brain-damaged are predominantly closed.

These differences in kinds of privateness are sufficiently pronounced to merit further consideration. The judgment of the privateness of a concept is based upon a criterion of relative inability of an observer to specify the limits of the conceptual grouping on the basis of a clearly denoted principle. Considered another way, a concept is private to the extent that the outsider would be unsure in anticipating whether additional objects would be included in or excluded from the sorter's conceptual grouping. While this criterion is applied in scoring both areas (open-private and closed-private), the quality of privateness in the two categories differs markedly.

An open-private concept has the quality of a

"personal idiom" (5). The most frequent types are idiosyncratic forms of symbolism, overabstractions, and those which show ellipses in the conceptual process. The several concepts which Cameron (3, 5) has introduced (e.g., overinclusiveness, asyndesis, metonymic distortion) to characterize schizophrenic thought are frequently apt descriptive terms for open-private conceptualization. They are so in the sense that they also refer to a form of thinking which lacks communicability, reality-testing, etc., but which involves the use of a relatively high-order of abstraction (e.g., "these are different forms of energy").

Schizophrenics, and others who employ an open-private form of thinking, frequently use some aspects of the objects in the sorting task simply to "trigger" a series of autistic associations. They are relatively insensitive to external "feedback," and they show a relative lack of self-monitoring, and lack of set toward communication. Little or none of this seems true of the thinking of those who employ predominantly closed-public concepts. Here, the quality of privateness in their conceptualization results not from *overabstraction*, but usually from *underabstraction*.

Many of the same subcategories are used in scoring both closed-private and open-private areas. The basic principle that distinguishes between these scoring areas is that of lower or higher order of conceptualization. Closed-private concepts are usually based upon the heterogeneous grouping of a number of objects "conceptualized" by the sorter at separate species-levels.⁷ Open-private concepts are usually based upon a synecdochal, or other type of loose-representational, relationship perceived by the sorter among the objects within the conceptual grouping. Thus, in an example of the "closed" narrative ("I could smoke this cigar after I finish my coffee"), each of the grouped objects (cigar, matches, spoon, sugar cubes) is appropriate in its *present species form* to the "principle" of belongingness expressed. On the other hand, in an example of the "open" narrative ("fellow might want to smoke after he eats his dinner and then go bowling"), the sugar-cubes are used to represent food in general; and, even more distantly, the attribute

of "greenness" of the paper square is abstracted and used synecdochally to represent the activity of bowling (inquiry revealed that the missing conceptual link was "bowling on the green").

The privateness of the "closed" narrative cited above lies in the failure of *S* to designate conceptual limits to his grouping; that is, it would be difficult for the outsider to predict whether or not each of a number of additional species-objects coming within the perceptual field of *S* would be included as "belonging" to the conceptual grouping, since no actual principle of collection is expressed. Thus, a closed-private concept is one in which there is virtually a complete *absence of a collecting principle* for a group of objects, considered by *S* to "belong together."

The virtual absence of a collecting, or generic, principle (closed-private conceptualization) and the use of conceptual classifications of a relatively low order of abstraction (closed-public conceptualization) are characteristic of the sorting behavior of the brain-damaged group. We have previously (15, 16) associated the terms "stimulus-boundness" and "reality-fixing" with this form of conceptual behavior. As distinguished from the concepts cited above to characterize schizophrenic thinking, these terms seem to us to bear a very close resemblance to some of the criteria invoked by Goldstein, Storch, Stern, and others to characterize concrete thinking in general; examples of these are impairment in discrimination between figure and ground (7), restriction of the conceptual response to the "immediate experience of the given thing or situation in its particular uniqueness" (7, p. 18), and conceptualization based upon a preponderance of concrete images (19). But it should be noted that these criteria for "concrete" behavior seem appropriate only for the specific kind of concreteness that is shown in a predominantly "closed" form of conceptualization.

On the basis of the findings of this study, the following conclusions are drawn:

1. The use of two isolated variables of conceptualization appears to be a more fruitful approach than the use of the dichotomous abstract-concrete formulation in predicting qualitative group differences in conceptualization.
2. The extreme differences in conceptual behavior between schizophrenic and brain-

⁷ The term "species" is defined as "a group of objects, each of which is 'called,' or usually identified, by the same name (for example, knives, pencils)" (15, p. 185).

damaged patients suggest that the two groups should not be represented as evidencing the same type of conceptual disorder.

SUMMARY

This was a study of differences in object-sorting behavior between schizophrenic and brain-damaged groups with respect to two conceptual variables—amount of social agreement and order of conceptual classification. It was hypothesized that the brain-damaged group shows a significantly lower order of conceptual classification than does the schizophrenic group.

The variable, social agreement, was scored in terms of the relative publicness-privateness of each conceptual sorting. The second variable, order of classification, was scored in terms of the number of attributes used in the definition. In addition to these two measures, five others were derived from the interaction of the two variables.

An analysis of covariance (controlling for intelligence) indicated that all six measures for which hypotheses were formulated yielded results in the predicted direction and that four of these mean differences were significant at the .01 level, or better. The greatest *F* ratio was for the closed-open variable, in keeping with the major hypothesis. Differences on all of the measures, except public-private, were greatest between the schizophrenic and brain-damaged groups with the results of the nonpsychiatric group falling in between.

REFERENCES

1. BOLLES, MARJORIE. The basis of pertinence. *Arch. Psychol.*, N. Y., 1937, No. 212.
2. BOLLES, MARJORIE, & GOLDSTEIN, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, 12, 42.
3. CAMERON, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, 50, No. 1 (Whole No. 221).
4. CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, 34, 265-270.
5. CAMERON, N. Experimental analysis of schizophrenic thinking. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944, Pp. 50-64.
6. CAPPS, H. M. Vocabulary changes in mental deterioration. *Arch. Psychol.*, N. Y., 1939, No. 242.
7. GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 17-40.
8. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2 (Whole No. 239).
9. GRASSI, J. R. The Fairfield Block Substitution Test for measuring intellectual impairment. *Psychiat. Quart.*, 1947, 21, 474-489.
10. GRASSI, J. R. *The Grassi Block Substitution Test for measuring organic brain pathology*. Springfield, Ill.: Charles C Thomas, 1953.
11. HANFMANN, EUGENIA, & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, No. 67.
12. HUNT, J. McV. Psychological loss in paretics and schizophrenics. *Amer. J. Psychol.*, 1935, 47, 458-463.
13. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
14. KASANIN, J. S. The disturbance of conceptual thinking in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 41-49.
15. MCGAUGHRAN, L. S. Predicting language behavior from object sorting. *J. abnorm. soc. Psychol.*, 1954, 49, 183-195.
16. MCGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and non-psychiatric groups. *J. abnorm. soc. Psychol.*, 1956, 52, 43-50.
17. MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, 67, No. 20. (Whole No. 370).
18. RAPAPORT, D. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.
19. STERN, W. *General psychology*. H. D. Spoerl (Trans.). New York: MacMillan, 1938.
20. STORCH, A. The primitive archaic forms of inner experience and thought in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1924, No. 36.
21. WERNER, H. *Comparative psychology of mental development*. E. B. Garside (Trans.). New York: Harper, 1940.
22. WHITE, W. A. The language of schizophrenia. *Res. Publ. Ass. Res. nerv. ment. Dis.*, 1928, 5, 323-343.
23. VIGOTSKY, I. S. Thought in schizophrenia. *Arch. Neurol. Psychiat.*, Chicago, 1934, 31, 1063-1077.
24. YATES, A. C. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, 51, 359-379.

Received February 20, 1956.

A COMPARISON BETWEEN IPSATIVE AND NORMATIVE RATINGS OF PERSONALITY¹

JACK BLOCK

Institute of Personality Assessment and Research, University of California, Berkeley

ALTHOUGH the distinction has been noted for some years (2, 5, 6), it was Cattell (3) who first named and paid detailed attention to the difference between "ipsative" and "normative" forms of psychological measurement. In ipsative measurement, the set of scores (traits, characteristics, behavioral measures) are weighted or ordered relative to that individual's own personal mean. In normative measurement, each score for an individual is evaluated relative to the mean score of a group of individuals. An illustration of the ipsative method is the rank ordering a subject may provide of his esthetic preferences among a variety of alternatives. Here the individual has a personal norm or indifference point in terms of which his choices may be evaluated. Normative measurement is exemplified by any test which employs group norms in terms of which an individual's score achieves meaning. As Cattell remarks, the distinction has considerable implication for the propriety of various methods of statistical evaluation, for the Q vs. R controversy in factor analysis, and for diverse other issues.

One of the contexts where the ipsative-normative distinction is important is where ratings of personality are employed. For reasons of convenience, necessity, and intrinsic interest, personality evaluations by observers have been and will continue to be widely used in research. Traditionally, personality ratings have been made normatively, e.g., 100 subjects are rated as to their "aggressiveness." More recently, the Q-sort method of personal-

ity description, an ipsative procedure, has achieved some usage.

There has been some question as to the relationship existing between the two measurement approaches. Cattell suggests that ipsative measurement treated normatively will provide results different from those derived by direct normative measurement (3, p. 296). More recently, Guilford has written, "Individual differences in ipsative measurements have little meaning because there is not a single scale for all individuals . . . (Ipsative scores) should not be used for correlation of variables over a population of individuals" (4, p. 528).

Clearly, there is good logical reason for suspecting a lack of equivalence between the two measurement approaches. The fact that in the ipsative case the mean in terms of which traits are evaluated may vary from individual to individual may make possible bizarre and inappropriate relationships.

However, the matter is open to empirical investigation as well as rational analysis. The present paper reports the results of two studies, in which ipsative ratings, treated normatively, were correlated with corresponding normative ratings, in a test of the functional equivalence of the two forms of measurement. The findings are, to the investigator at least, somewhat surprising. A number of current rating and assessment procedures perhaps bear revision in the light of these data.

ANALYSIS I

¹ This research was supported in part by the United States Air Force under Contract No. AF 18 (600)-8, monitored by Technical Director, Detachment #7 (Officer Education Research Laboratory), Air Force Personnel and Training Research Center, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

Personal views or opinions expressed or implied in this publication are not to be construed as necessarily carrying the official sanction of the Department of the Air Force or of the Air Research and Development Command.

As part of an assessment program at the Institute of Personality Assessment and Research, 100 Air Force officers were observed for a three-day period by a staff of eight psychologists. At the end of assessment, each of the ten staff members rated each subject (S) on 30 aspects of personality. Each rater used a 5-point rating scale and, in evaluating the 100 Ss, considered each rating variable singly. The ratings were afterward averaged to form for each S one score for each dimension. These data served as the normative measures in the comparison.

Eight of the ten staff members also described each subject by means of a 76-item Q-sort deck. The items were evaluated vis-à-vis each other for relevance and salience in characterizing each S's personality, and sorted along a 9-point continuum. The scale scores for

TABLE 1

CORRELATIONS BETWEEN NORMATIVE RATINGS AND EQUIVALENT Q-SORT ITEMS: FIRST STUDY
(Reliability of normative and ipsative ratings are in parentheses)

	Ipsative-Normative Correlation	Corrected Ipsative-Normative Correlation
1. "Ability to communicate" (.90) vs. "Communicates ideas clearly and effectively" (.78)	.88	1.05
2. "Adaptive flexibility" (.87) vs. "Is rigid; inflexible in thought and action" (.80)	-.78	.94
3. "Breadth of interests" (.88) vs. "Has a narrow range of interests" (.91)	-.85	.96
4. "Constriction" (.83) vs. "Overcontrols his impulses; is inhibited; needlessly delays or denies gratification" (.88)	.82	.96
5. "Dominance" (.90) vs. "Takes an ascendant role in his relations with others" (.90)	.82	.91
6. "Evaluation of ideas" (.80) vs. "Has the ability to draw dependable and practical inferences; common sense; good judgment" (.73)	.73	.96
7. "Fairmindedness" (.81) vs. "Allows personal bias, spite or dogmatism to enter into his judgment of issues" (.69)	-.70	.93
8. "Fluency of ideas" (.90) vs. "Is stereotyped and unoriginal in his approach to problems" (.84)	-.84	.97
9. "Good judgment" (.88) vs. "Has the ability to draw dependable and practical inferences; common sense; good judgment" (.73)	.75	.94
10. "Health and vitality" (.79) vs. "Is active and vigorous" (.84)	.71	.88
11. "Impulsivity" (.83) vs. "Undercontrols his impulses; acts with insufficient thinking and deliberation; unable to delay gratification" (.75)	.69	.87
12. "Intellectual competence" (.90) vs. "Has a high degree of intellectual ability" (.82)	.86	1.00
13. "Leadership" (.94) vs. "Is an effective leader" (.80)	.79	.91
14. "Likeability" (.73) vs. "Is a likeable person (N.B. the subject's general acceptability rather than the rater's personal reactions is intended.)" (.81)	.66	.86
15. "Masculinity" (.86) vs. "Is masculine in his style and manner of behavior" (.83)	.79	.94
16. "Military & social presence" (.79) vs. "Lacks social poise and presence; becomes rattled and upset in social situations" (.77)	-.79	1.01
17. "Naturalness" (.74) vs. "Is natural; free from pretense; unaffected" (.80)	.73	.95
18. "Originality" (.86) vs. "Is stereotyped and unoriginal in his approach to problems" (.84)	-.87	1.02
19. "Positive valuation of the military identity" (.79) vs. "Identifies with values, actions, attitudes, etc., of his professional group" (.72)	.62	.83
20. "Rigidity" (.79) vs. "Is rigid; inflexible in thought and action" (.80)	.77	.97
21. "Self-confidence" (.80) vs. "Lacks confidence in his own ability" (.82)	-.80	.99
22. "Self-defensiveness" (.71) vs. "Is self-defensive; rationalizes, excuses, blames" (.64)	.63	.94
23. "Self-insight" (.70) vs. "Lacks insight into his own motives and behavior" (.61)	-.66	1.02
24. "Verbal fluency" (.90) vs. "Is verbally fluent; conversationally facile" (.83)	.84	.98

each item were afterward summed across raters in order to generate for each *S* one score for each *Q* item. These data served as the ipsative measures in the comparison.

It proved possible to match 24 of the 30 normative ratings with *Q* items. Matching was done on an *a priori* basis by two psychologists and the criterion for matching was simply that the sensible meaning of the *Q* item be considered to correspond fairly closely, if not exactly, to the definition of the normative ratings.

For each of the 24 pairings, the correlation between normative and ipsative scores was computed. These are listed in Table 1, which also identifies the rating and *Q*-item variables, the respective split-half reliabilities, and the ipsative-normative correlations corrected for attenuation. Over all comparisons, the median ipsative-normative correlation, corrected for attenuation, is .95.

This figure makes no allowance for the lack of exact equivalence in certain of the comparisons.

ANALYSIS II

The second comparison employed data provided by a one-day assessment of 70 applicants to medical school. A group of seven psychologists contributed normative ratings on 25 variables while a *separate* group of seven psychologists *Q* sorted the *Ss* using a 108-item "clinical" *Q* deck. This deck was oriented toward more highly inferential ("genotypical") features of personality than the descriptive ("phenotypical") items used in the first study, or the normatively rated variables employed in both analyses. Because of the greater complexity of these *Q* items, the matching of normative ratings with ipsative items proved to be more difficult and in certain instances rather arbitrary.

TABLE 2

CORRELATIONS BETWEEN NORMATIVE RATINGS AND EQUIVALENT Q-SORT ITEMS: SECOND STUDY
(Reliabilities of normative and ipsative ratings are in parentheses)

	Ipsative-Normative Correlation	Corrected Ipsative-Normative Correlation
1. "Breadth of interests" (.83) vs. "Has a wide range of interests" (.58)	.74	1.07
2. "Dominance" (.86) vs. "Behaves in an assertive fashion" (.82)	.53	.63
3. "Ego-strength" (.84) vs. "Has a brittle ego-defense system; has a small margin of integration; precarious self-control" (.61)	-.44	-.62
4. "Impulsivity" (.75) vs. "His various needs tend toward relatively direct and uncontrolled expression; unable to delay gratification; acts out" (.59)	.64	.95
5. "Intellectual competence" (.76) vs. "Has a high degree of intellectual competence" (.69)	.55	.76
6. "Introspectiveness" (.85) vs. "Is introspective and concerned with self as an object; frequently self-aware" (.71)	.60	.77
7. "Masculinity" (.82) vs. "Is masculine in his style and manner of behavior" (.68)	.65	.86
8. "Need for achievement" (.66) vs. "Has high aspiration level for self; ambitious; wants to get ahead" (.45)	.31	.57
9. "Originality" (.69) vs. "Is original and imaginative" (N.B. <i>quality</i> of thought rather than quantity is to be emphasized) (.69)	.61	.89
10. "Responsibility" (.68) vs. "Is a dependable and responsible person" (.46)	.54	.96
11. "Self-insight" (.82) vs. "Has insight into his own motives and behavior" (.49)	.48	.75
12. "Social acuity" (.68) vs. "Is socially perceptive; responsive to a wide range of social cues" (.65)	.52	.78
13. "Social presence" (.84) vs. "Has social poise and presence; appears socially at ease" (.82)	.58	.69
14. "Verbal fluency" (.72) vs. "Is a talkative individual" (.78)	.61	.81

Fourteen matchings were deemed possible. Ipsative-normative correlations then were computed from scores derived as in the first analysis. These are presented in Table 2, along with the split-half reliabilities of the ipsative and normative measures, and the ipsative-normative correlations corrected for attenuation.

The findings from the second analysis are not quite so spectacular as in the first comparison. But the median ipsative-normative correlation, corrected for attenuation, is still .79. This is quite a respectable figure and is found despite the fact that the second analysis provides a much stricter and perhaps unfair test of the equivalence between ipsative and normative ratings. By using separate groups of raters in the second study, the possibility is eliminated of correlated rating error, a factor which from one point of view may have affected the results of the first analysis. But the use of separate rater groups introduces a source of variance which is extraneous to the comparison of interest, i.e., the results provided by the two methods. Because the correlation between the two measurement approaches in the second study reflects possible real differences between the two sets of raters as well as a possible real difference between the two methods, it is reasonable to find the coefficient somewhat lower.

Another cause of attenuation is to be found in the poorer ipsative-normative matchings in the second study. The ipsative ratings in the second analysis were more inferential and depended less on behavior than did the normative ratings.² Because the matchings were

not so complete as in the first study, the ipsative-normative correlations should show some decrease.

Taking the two studies together, therefore, it would appear that there is an almost complete functional identity between the results obtained from ipsative ratings treated normatively and conventionally acquired normative rating data. The one can be directly substituted for the other, with insignificant loss of discriminations.

DISCUSSION

Why is it that the ipsative-normative ratings appear to be so fully equivalent? One possibility would seem to lie in the nature of the ipsative judgmental process. In ipsative ratings, the decision between qualitatively different variables as to "salience" is made with reference to a very large apperceptive body of information and evaluation as to the psychology of human beings. This "adaptation level" is

from the sets of reliability estimates of these ratings. In the first study, ipsative and normative dimensions required approximately the same level of rater inference and their reliabilities were about equal (means of .80 and .84, respectively, computed via the *Z* transformation). But in the second study, ipsative ratings were clearly more genotypic in their emphasis. Here the ipsative reliabilities suffered where the normative reliabilities remained the same (means of .66 and .78 respectively; computed by the *Z* method.)

² Suggestive evidence for order of complexity as affecting the ipsative-normative correlations comes

based upon some sum or integration of the rater's lifetime experience with and conceptualization of various personalities. The frame of reference from which the ipsative judgment is made is therefore in actuality normatively derived, and it is not surprising (with after-the-fact wisdom) that the considerable equivalence exists.

But the normative reference base from which ipsative ratings are made is a much broader one than that employed for most normative ratings. In the case of ipsative ratings, the normative adaptation-level is a very stable one, accumulated over many years and not subject to radical change when a short run of disproportionate personalities is encountered. Something like an absolute scaling of the rater's impressions is obtained.

In the case of directly normative ratings, the reference frame is usually a specifically delimited population; e.g., military officers, applicants to medical school, and so on. Because in an absolute sense populations may reliably differ with respect to their mean and range on various personality variables, relationships among these variables are likely to shift appreciably as a function of the particular samples studied (1). A "masculine" graduate student may not be a "masculine" military officer when rated vis-à-vis this second population. But in the generalized normative frame in terms of which the ipsative rating is made, his rating for "masculinity" is likely to remain the same, regardless of the particular sample of which he is a member. Ipsative ratings, normatively treated, thus offer the possibility of providing more stable interrelationships among variables than has usually been the case.

Which method, ipsative or normative, is to be preferred in research settings? The normative method clearly has a more immediate convenience since only one variable is dealt with at one time and the subject sample may be readily evaluated in terms of the particular one or two dimensions of interest. To achieve ipsative scores on but a few dimensions, *all* of the variables in the ipsative sample of traits must be evaluated for *each* subject, a task which may be quite formidable. It should be noted, though, that the normative approach loses its advantage of convenience as the number of variables of interest increases.

Although there are many contexts where the

ipsative method is inefficient or inappropriate, there is one somewhat unusual benefit to be enjoyed from its use. Frequently, in research retrospect, e.g., in the summing up of longitudinal studies, one wishes that an additional variable or conception of personality had been evaluated with reference to the now departed subject sample. With the normative method, there is no simple or easily justified way in which these belated evaluations may be gathered. Memories prove faulty or the originally participating raters may no longer be available. With ipsative data, however, the desired ratings are quickly achieved.

As an empirical illustration of how the omission may be rectified, consider the variable "Over-all Military Effectiveness in a Command Position," normatively rated in the first study reported above. This variable could not be matched with any of the 76 ipsative Q items and was not employed in the main body of the analyses. Let us suppose that some years after the actual study, some evaluation of this dimension was desired for criterion purposes. An ipsative measure could be obtained in the following manner:

1. Each of a number of staff psychologists would Q-sort their conception of the Air Force officer manifesting the greatest degree of "Over-all Military Effectiveness in a Command Position." In actual fact, the average intercorrelation of five conceptual Q sorts was .80. Only two of the five sorters of the concept had previously observed the officer sample and their sorts did not interrelate differently than did the remaining three.

2. A composite of this conceptual model would then be constructed by summing the Q-item placements across sorters. In the actual case, this composite has a reliability of .89.

3. The empirical Q sorts of each of the 100 officers would then be correlated with the conceptual composite.

4. These correlations with the composite could be treated as a set of scores ordering the subjects on the specified dimension. In the actual illustration, when these correlations, considered as scores, were correlated with the initially obtained normative ratings of the dimension, the uncorrected correlation was .85. Substantially the same degree of equivalence existed in several other tests of this technique.

Thus it would appear that by use of conceptual Q sorts, and an original Q set which is comprehensive in nature, scores can be derived from actual Q sorts which pertain to dimensions not specifically included in the Q set or otherwise anticipated. The normative method does not offer this useful facility in any convenient form, in part because normative ratings are usually too few in number, but most particularly because normative ratings as usually gathered are too tied to the particular referent population.

It is perhaps wise to close on a cautionary note. The present paper deals only with the comparison of ipsative and normative ratings. The ipsative-normative distinction has ramifications much beyond those touched upon here. In particular, the issues involved in comparing ipsative and normative *objective measures* await empirical study.

SUMMARY

Ipsative and normative ratings of personality were compared in two separate studies. Both analyses showed an almost complete equivalence between the two methods. A possible reason for this relationship was discussed and some of the advantages of the ipsative approach presented.

REFERENCES

1. BLOCK, J. The difference between Q and R. *Psychol. Rev.*, 1955, **62**, 356-358.
2. BURT, C. Correlations between persons. *Brit. J. Psychol.*, 1937, **28**, 59-96.
3. CATTELL, R. B. Psychological measurement: ipsative, normative, and interactive. *Psychol. Rev.*, 1944, **51**, 292-303.
4. GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
5. STEPHENSON, W. Correlating persons instead of tests. *Charact. & Pers.*, 1935, **4**, 17-24.
6. THOMSON, G. *The factorial analysis of human ability*. London: Univer. of London Press, 1939.

Received February 21, 1956.

THE EFFECTS OF GROUP EXPECTATIONS AND SELF-ESTEEM UPON SELF-EVALUATION¹

EZRA STOTLAND, STANLEY THORLEY, EDWIN THOMAS, ARTHUR R. COHEN,²
AND ALVIN ZANDER

University of Michigan

RESEARCH on the level of aspiration has demonstrated quite well that a person's evaluation of his performance does not depend upon his absolute level of achievement alone. What counts is the level of achievement relative to his aspiration—his goal line. If the achievement lies on or above the goal line he will probably rate his performance as good, depending upon the size of this difference. These studies have been carefully reviewed and interpreted by Lewin *et al.* (5).

The placement of the goal line, whether it is high or low, may be influenced by different scales of reference. Two kinds of reference scales have been mentioned often. One is a source of pressure from other persons or groups to perform in a particular fashion on that activity. The other involves pressure from the person's own expectations concerning his likely performance in that activity, arising from past experiences. In the present study we examine the effect of a specific level of achievement upon an individual's evaluations of his performance when the achievement is relative to a level of aspiration established by a group, and when the achievement is also relative to the member's stabilized expectations about himself as represented by his self-esteem.

Group Expectations

The ability of a group to determine levels of aspiration for individuals is discussed by Lewin *et al.* (5). Several different approaches have been used in studying these events. In the method most close to the purposes of this study, a group's level of aspiration was described to the members and they then accepted this level as their own (2).

This approach suggests that we may conceive of the level of aspiration as a group

standard, which is the resultant of pressures to achieve that level exerted by the group upon the members. The power of a group is its ability to influence the beliefs or behavior of its members. If the group has enough power to influence the members, and the group standard is high, the members will set high levels of aspiration for themselves. If the standard is low, they will set low levels of personal aspiration (6).

The power of a group to influence its members may vary under different conditions. It is evident in the work of Festinger (3) that a group's power is greater the more that the group is attractive to its members. Thus, the members' levels of aspiration will be more closely modeled after the aspiration of the group, the more that the group is attractive to them. If attraction to the group is constant and moderately high among all members, however, another matter determines whether they conform to their group's level of aspiration. This is the relevance of the activity.

Relevance refers to the degree to which an activity is important to the purposes of, or for the maintenance of, a group. Schachter (7) has shown that the power of the group is likely to be stronger upon persons engaging in an activity relevant to the group than upon persons performing a nonrelevant activity. It follows, that the aspiration level set by a group concerning a relevant task tends to be supported by stronger pressures on the person to perform up to that level, than does an aspiration level set for a nonrelevant task.

Regardless of the relevance of an activity, some persons may accept a group aspiration as their own more than do other individuals. We will not occupy ourselves here with the reasons for variation in accepting the group's level of aspiration. It is expected, however, that the group's level of aspiration would have more effect upon personal evaluation of a success or a failure, the more the members accept the level of aspiration as their own.

For many groups in our society there are

¹ The work on this study was financed by a grant-in-aid [M-701 (C)] from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, Alvin Zander, principal investigator.

² Arthur R. Cohen is now at Yale University.

external criteria which help a member to judge how well he has performed an activity which is of interest to the group. An athlete knows whether he has won, and a salesman knows whether he has made the sale. The team's, or staff's, level of aspiration is not needed for matters such as these to help a person know if he has succeeded or failed, but the group's aspiration is pertinent in influencing the extent to which the individual evaluates an objective failure as a failure and an objective success as a success, assuming, of course, he expects that the group will become aware of his performance. In the present study the subjects were made either to fail or to succeed, and had a clear knowledge of the level of their performance. The major issue is: what conditions in the group may influence a person's evaluation of his performance when he knows that he has succeeded or failed?

From the considerations thus far, two general propositions are derived: (a) When a task is relevant to a group, the group's level of aspiration will raise or lower a member's evaluation of his performance more, after an experience of objective success or failure, than when the task is nonrelevant to the group. (b) The group's level of aspiration will raise or lower a member's evaluation of his performance to the degree that he accepts the group's level of aspiration as his own.

Specific hypotheses suggested by these propositions, but more amenable to verification, are as follows:

1. A level of aspiration set by a group for its members is more often accepted by a member as a personal goal when the activity is relevant to the group than when it is nonrelevant.

2. A failure on a relevant task generates poorer self-evaluation of a person's performance if a group's aspirations are high than if they are low (while the evaluation of a failure on a nonrelevant task is not affected by different levels of the group's aspirations).

2-a. A failure when a person accepts the group's level of aspiration as his own, generates lower self-evaluation of his performance if the group's aspirations are high than if they are low (while evaluation of a failure, when the group's aspiration is rejected, is not affected by different levels of the group's aspirations).

3. A success on a relevant task generates better self-evaluation of a person's performance if a group's aspirations are high than if these aspirations are low (while the evaluation of a success on a nonrelevant task is not affected by different levels of the group's aspirations).

3-a. A success when a person accepts the group's level of aspiration as his own, generates better self-evaluation of his performance if the group's aspirations are high than if they are low (while the evaluation of a success when the group's aspiration is rejected is not affected by different levels of the group's aspirations).

Self-Esteem

When a person evaluates his level of achievement he refers not only to the expectations of other persons but to the concept he has of himself as well. The degree of an individual's self-esteem, therefore, is expected to affect the way in which an individual evaluates his performance in a particular situation. *Self-esteem* is a function of the degree of coincidence between an individual's ideal and actual concepts of himself. The more that the actual and ideal concepts coincide, the higher the self-esteem.

Persons who are high in self-esteem, we assume, differ from those who are low in the way they react to an immediate experience which has been a test of their abilities. Cohen (1) has reported that persons with high self-esteem, as compared to those with low, are characterized by tendencies to protect themselves from negative self-evaluation. It is not clear why this should occur. Perhaps persons with high self-esteem have learned to overlook their objective failures and to concentrate upon their successes in building their concept of themselves, or perhaps they have consistently performed in accordance with their ideals and hence are less threatened by an experience of failure. In any event, it would be expected that persons with high self-esteem would be able to evaluate an objective failure as a small failure and an objective success as a large success. Persons with low self-esteem, however, because they do not protect themselves from negative evaluation, are more likely to evaluate an objective failure as a very poor performance and a success as a small success.

When the person is a member of a group which holds certain standards for his performance, how does the level of his self-esteem affect his evaluation of his achievement?

A number of different answers to this question appear plausible. A choice among the alternatives cannot be made with confidence, however, because there is no way of anticipating whether the potency of a scale of reference provided by the power of the group will equal that offered by a person's self-esteem; or whether the weight of the group and individual influences will differ under certain conditions, but not under others. Speculation about these matters is postponed until it is clear how the group standards and self-esteem interact in determining the way a person evaluates his level of achievement.

METHOD

An opportunity to examine these phenomena was created in a laboratory experiment in which the effects of both group pressures and self-esteem upon self-evaluation could be observed. Two levels of group aspiration were created, termed high and low expectations hereafter. At the same time, the relevance of the activity to the group's purpose was varied by describing the S's task as essential to the group goal in one set of conditions and nonrelevant to the group goal in the other. Thus, there were four conditions consisting of high or low expectations in combination with relevant or nonrelevant tasks. Finally, within each of these conditions, half of the Ss were made to fail on their tasks and half to succeed. The dependent variables were measured in a questionnaire administered after the Ss had worked on the tasks but while they were assuming that the experiment was not yet finished. The variables were the Ss' evaluations of their performances on the task and their concern with the group's expectations (i.e., acceptance of the group's level of expectations). All were measured on 6-point graphic rating scales. In addition, a measure of self-esteem was administered before the experiment was started, and again as part of the second questionnaire.

Measurement of self-esteem. The measure of self-esteem was designed in accord with the definition of the concept and modeled after an instrument used by Cohen (1). A person's score describes the degree to which he meets his needs in ways he considers to be ideal.

The items of the test area series of brief paragraphs each describing a hypothetical situation in which a person faces a potential frustration of a need. Problems of meeting needs in five representative areas are included: achievement, recognition, affiliation, cognition, and autonomy. After reading each paragraph, the S indicates how a person would react by choosing from among five behavioral alternatives. He first chooses the reactions *most* and *least* descriptive of the way a person should ideally behave. Then, after a second read-

ing of each paragraph, he chooses the reactions *most* and *least* descriptive of the way a person actually would behave.

There follows an example of one of the hypothetical situations and the alternative responses for it.

A person has always valued his sense of freedom and independence. Because of what others think and say, this sense of freedom and independence is greatly hampered.

Most Least

a. Reason that maybe a great many mutual responsibilities with others are most important after all. . . .

b. Figure that while things are confining now, knowing where one stands in relation to others is half the battle. . . .

c. Tend to turn his thoughts to more pleasant aspects of his dealings with others. . . .

d. Face the fact that other people are over-eager to maintain their opinions and ideas in the face of his. . . .

e. Think twice before overly involving himself with others because these experiences are uncomfortable for him. . . .

The responses are scored by giving a +1 for each occasion where the actual most preferred reaction is the same as the ideal most preferred reaction, and where the actual least preferred is the same as the least preferred. A score of -1 is given for a coincidence of the actual most preferred with the ideal least preferred, and of the actual least preferred reaction with the ideal most preferred. If no coincidence occurs, no score is given. The person's score is obtained by summing across all items.

In previous work with a longer version of this measure on a sample of sixty persons, the odd-even split-half correlation was .84 (.91 when corrected by the Spearman-Brown formula). In order to measure self-esteem in the laboratory, however, the requirements made it necessary to use only one-third of the hypothetical situations of the original instrument. Scores on all the items were significantly related to total scores in the larger instrument.

The sequential phases of the experiment are now described.

Group formation on a cooperative task. The purposes of the opening part of the experiment were, first, to form a basis for the subsequent manipulation of relevance and, second, to make membership in the group salient for the Ss. Making membership salient, it was assumed, made attraction to the group high so that group pressures were acceptable to the members. These purposes were accomplished by having the group work and succeed on a cooperative task, in competition with other, fictitious groups.

The Ss were 175 male undergraduates at the University of Michigan who volunteered to participate in the study. They came to the laboratory four, five, or six at each session, and care was used to ensure that they were strangers to one another. As they came into the laboratory, they were taken to individual cubicles to fill out the self-esteem questionnaire. When all of them had finished, they were seated in a circle to receive instructions, and were told that they would be working

on a cooperative task in competition with other teams. In order to help create a "team spirit," they were asked to pick a name for themselves.

The group task given them was to form two letters of the alphabet. The letter *U* was first to be formed in a groove in a board by filling it with various pieces of different sizes and shapes, as in a jigsaw puzzle. After each *S* had been given 12 pieces, the board was passed from person to person around a circular table 12 times. Each man placed one piece in the groove each turn, flush against the pieces already there. Thus, each person made an equal contribution to achievement of the group goal: to complete the *U* as fast as possible.

When the group was finished, the *E* announced the time and publicly placed a pin for the team on a graph on which the speeds of the "other" groups were indicated. This graph was arranged so that each group discovered it had done moderately well relative to others. Moderate achievement was stressed in order to prevent strong feelings of either success or failure which could confound the effect of later procedures.

Relevance and nonrelevance inductions. After the group had completed its *U*, the *Ss* were given tasks to be done in private, with different instructions for the relevant and nonrelevant conditions. When the individual task was made relevant, the *E* told them that the second letter would later be formed by the group from parts each member would provide. The one piece in his individual puzzle which fitted within the design, but did not touch any of the sides, was the piece required by the group. The *Ss* were then told to return to their cubicles to work on the individual puzzles. They were allowed 12 minutes for the task.

In the nonrelevant condition, the *Ss* were asked to work on the same task, but were told that it had nothing at all to do with the group goal and that they would each receive individual scores on whether or not they were able to complete the task. They were informed that they would return to the second group letter at a later time.

Expectations induction. The two levels of expectations, high and low, were induced as follows. After explaining the individual tasks, and before the *Ss* began to work on them the *E* said:

"We often find it helpful in group situations like this to give each person a clear idea of how well his group expects him to perform. It might be that the group expects him to meet a given level of performance, or it might be that the group feels it cannot really expect a member to do that well on a difficult task. I would like you to decide which of these expectations you want to impose on yourselves. I am going to pass out these ballots on which you can indicate which of these expectations you want. Please read the descriptions of the expectations very carefully before voting. I will collect and count the ballots."

The questions read as follows:

Please check one

1. *High expectations:* I vote that each of us will be expected by the others to complete his individual task, and that we will expect the others to do the same.

2. *Low expectations:* I vote that each of us will be expected to do only as well as he can toward com-

pleting his individual task, and that we expect the others to do the same.

The two expectations conditions were created by giving the *Ss* contrived tabulations favoring high or low levels.

Achievement level. In the success condition, the task was easy enough to allow 95 per cent of the *Ss* to complete it in the allotted time. In the failure condition, the *Ss* were given puzzles which were impossible to complete because no piece met the criterion of a successful solution. Blank jigsaw puzzles (i.e., no pictures on the faces of the pieces) were used in order to make it uncertain as to how close they had come to the correct way of assembling the parts if they did not complete it. Thus a failure could only be an objective failure in fact, and no gradations of partial success were possible.

Questionnaires. After the *Ss* were through working on the individual task, they were again given a questionnaire with the understanding that they would do the second group letter after the form was finished. When the questionnaire was completed, they were told that there was no second group task and the experiment was over. The purpose of the study was then explained, and all questions answered.

RESULTS³

Group expectations and self-evaluation. The first hypothesis, which asserts that members will accept the level of the group's expectations as a personal goal line more often when the task is relevant than when it is nonrelevant, was tested by examining the degree to which the *Ss* said they were "concerned" about the team's expectations while they had been working on their individual tasks. The mean concern in the relevant condition is 4.00, which is significantly greater than the mean of 3.27 in the nonrelevant condition ($p = .01$, by t test). The data support the hypothesis.

The second hypothesis states that a person who fails on a task which is relevant to the group will evaluate his performance lower, if the group's expectations are high than if the group's expectations are low. In Table 1 it may be seen that in the relevant condition the mean self-evaluation is lower (2.05) when the group expectations are high, than when the expectations are low (2.65). The difference between these means is significant by a one-tailed t test at the .05 level. When the task was nonrelevant, and the members failed, the

³ A comparison of any two means which is not specifically mentioned in the text may be assumed by the reader to be not significant at the .05 level or better, except for the contrasts between means of self-evaluation in success conditions with those in failure conditions, all of which are significant.

level of the group's expectations creates no significant difference between the mean evaluations of performance (2.00 and 1.90). The data support the second hypothesis.

Hypothesis 2-a states that when persons fail on an activity, they will evaluate their achievement as poorer when the group's expectations are high, and will evaluate their performance as better when the group expectations are low. This effect will be more marked for persons with high concern than for those with little concern about the group's expectations.

In Table 2 the mean self-evaluation for those Ss with high concern for the group's expectations is significantly lower, where the group's expectations are high, for a relevant task which is failed (1.83), than where the group's expectations are low (2.83). This difference is significant by a *t* test at .05. The persons who were little concerned about the group's expectations do not differ significantly in their evaluation of their failure experience according to whether the group's expectations were high or low (2.36 and 2.20 respectively). When the task was nonrelevant, the members' degree of concern about the group's expectations does not generate a significant difference in the way that they evaluate a failure. It appears that the group's level of expectations had its strongest effect upon self-ratings of failure when the task was relevant and the members were highly concerned about the group's expectations. These results support Hypothesis 2-a.

The third hypothesis proposes that a person who succeeds on a relevant task will evaluate his performance lower when group expectations are low, and higher when group expectations are high. No significant differences are found (Table 1) in comparing the means of those who

TABLE 1
MEAN SELF-EVALUATION OF SUCCESS ON TASK FOR ALL
CONDITIONS
(*N*'s in parentheses)

Achievement level	Relevant task		Nonrelevant task	
	High group expectations	Low group expectations	High group expectations	Low group expectations
Failure	2.05 (25)	2.65 (21)	2.00 (20)	1.90 (25)
Success	4.80 (22)	4.59 (22)	4.95 (20)	4.80 (20)

TABLE 2

MEAN SELF-EVALUATION OF SUCCESS ON TASK FOR ALL
CONDITIONS, COMPARING PERSONS WITH HIGH AND
LOW ACCEPTANCE OF GROUP'S EXPECTATIONS
(*N*'s in parentheses)

Achievement level	Relevant task		Nonrelevant task	
	High group expectations	Low group expectations	High group expectations	Low group expectations
A. High acceptance of group's expectations				
Failure	1.83 (12)	2.83 (12)	1.80 (10)	2.40 (5)
Success	4.73 (15)	4.69 (13)	5.09 (11)	5.00 (9)
B. Low acceptance of group's expectations				
Failure	2.36 (11)	2.20 (10)	2.60 (10)	1.69 (16)
Success	4.50 (10)	4.60 (10)	4.78 (9)	4.69 (16)

succeeded, regardless of the group's level of aspiration or the relevance of the task. Thus, the third hypothesis is not supported. In Hypothesis 3-a it was predicted that a self-rating of success after a success experience would be a direct function of the group's level of expectations, more so for persons with high concern than for those with little concern. The anticipated results did not occur; thus Hypothesis 3-a is not supported.

Self-esteem and self-evaluation. We examine now the effect of an individual's self-esteem upon his evaluation of his achievement. Self-esteem was considered to be an enduring personality characteristic which the Ss brought with them to the laboratory. It was assumed that it would not be changed by the experience there. The reasonableness of this assumption is shown by a comparison of the mean self-esteem scores obtained prior to the experiment with scores obtained at the time of the final questionnaire. No significant differences appeared between any of the experimental conditions. Thus, the events in the experiment did not change the participant's self-esteem to a significant degree.

Persons with high self-esteem have little more positive evaluation of their own performance than those with low self-esteem, when all other matters are equal. The mean self-ratings made by persons with high self-esteem is 3.95 and the mean for those with low self-esteem is 3.88.

Members who failed on the task, however, who were high in self-esteem evaluated their

TABLE 3
MEAN SELF-EVALUATION OF SUCCESS ON TASK FOR ALL
CONDITIONS, COMPARING PERSONS WITH HIGH
AND LOW SELF-ESTEEM
(*N*'s in parentheses)

Achievement level	Relevant task		Nonrelevant task	
	High group expectations	Low group expectations	High group expectations	Low group expectations
A. Persons with high self-esteem				
Failure	2.36 (8)	2.60 (15)	2.67 (12)	2.45 (11)
Success	4.70 (10)	4.42 (12)	4.82 (11)	5.25 (8)
B. Persons with low self-esteem				
Failure	1.86 (14)	2.43 (7)	1.50 (8)	1.20 (10)
Success	4.87 (15)	4.78 (9)	5.11 (9)	4.59 (17)

performance higher than those with low self-esteem, all other variables constant. The means are 2.55 and 1.75 respectively, and are significantly different at the .05 level. Among the persons who succeeded, those who were high in self-esteem were not at all different from persons low in self-esteem in the self-ratings they assigned to their achievement (4.75 and 4.83 respectively). The amount of the member's self-esteem, then, affected evaluation of performance when the person failed, but not when he succeeded.

It was apparent earlier that the members were more likely to evaluate a failure in accordance with the group's level of expectations when the task was relevant than when it was nonrelevant. It is striking, now, to observe that self-esteem appears to have served as a scheme of reference for the members more in the nonrelevant condition than in the relevant one. In Table 3 it may be seen that where the expectations of the group are *high* and the person fails on a task which is nonrelevant, those with high self-esteem have higher evaluations (2.67), than the persons with low self-esteem (1.50). The difference between these means is significant at the .05 level. Where the expectations of the group are *low* and the persons fail on a task which is nonrelevant, those with high self-esteem again have higher self-evaluation (2.45) than the individuals with low self-esteem (1.20). The difference between these means is significant at the .01 level. When the *Ss* failed on a relevant task, no significant differences were found in the self-ratings of those with high self-esteem

compared to those with low self-esteem. It appears that the members were more likely to evaluate their performance according to their levels of self-esteem when the power of the group was weakest (i.e., the task was non-relevant), but less likely to do so when the power was strongest (i.e., the task was relevant). This statement is true, however, only when the members failed. Among those who succeeded on the task, persons with high self-esteem had no different feelings of success than those with low self-esteem when level of expectation was held constant, regardless of the relevance of the task.

In the opening discussion it was anticipated that persons with high self-esteem would better protect themselves against poor evaluations than those with low self-esteem. We have noted that persons with high self-esteem compared to ones with low self-esteem, evaluate their performances more favorably when they have failed. We have also noted that when self-esteem alone is varied there is no difference in the evaluation of their performance. These last data suggest that persons with high self-esteem do not always rate their performance more highly than persons with low self-esteem. Can we say, then, that persons with high self-esteem appear to protect themselves? Do they do this regardless of the group's expectations? The data in Table 4 are pertinent to these questions. They indicate that persons with high self-esteem, compared to individuals with low self-esteem, are cautious about becoming concerned with the group's expectations.

The degree of concern felt for the group's

TABLE 4
MEAN CONCERN WITH GROUP EXPECTATIONS FOR ALL
CONDITIONS, COMPARING PERSONS WITH HIGH
AND LOW SELF-ESTEEM
(*N*'s in Parentheses)

Achievement level	Relevant task		Nonrelevant task	
	High group expectations	Low group expectations	High group expectations	Low group expectations
A. Persons with high self-esteem				
Failure	4.25 (8)	4.20 (15)	2.83 (12)	2.91 (11)
Success	3.40 (10)	3.72 (13)	3.18 (11)	4.11 (9)
B. Persons with low self-esteem				
Failure	3.86 (14)	3.69 (7)	4.63 (8)	2.70 (10)
Success	4.67 (15)	3.63 (8)	3.67 (9)	2.94 (17)

expectations differed for persons with high self-esteem compared to those with low esteem, on the nonrelevant tasks. As may be seen in Table 4, persons with high self-esteem stated that they had little concern for the group's expectations when they had failed and the group's expectations were high (2.83). Those with low self-esteem, on the other hand, when they had failed under high group expectations, expressed high concern for the group's expectations (4.63). The difference between these means is significant at the .01 level. In addition, participants with high self-esteem said they were most concerned about the group's expectations when they had succeeded and the level of the group's expectations was low (4.11). While Ss with low self-esteem were little concerned with the group's expectations when they had succeeded and the group's expectations were low (2.94). The difference between these means is significant, however, at only the .10 level.

It appears that Ss with high self-esteem became interested in the group's expectations when the expectation levels were easy to achieve and in fact had been reached, and were less interested in the group's expectations when these expectations were difficult to reach and they had failed to perform up to that level. Persons with low self-esteem were in contrast: they became attentive to the group's wishes when these were difficult to gain and the person had failed, while they were less attentive to the expectations when the level was easily available and they had already succeeded. The results suggest that persons with high self-esteem better protect themselves against unfavorable evaluation by avoiding involvement in the group's expectations when an unfavorable comparison with others would be likely to occur.

In the relevant situation the member's self-esteem was not related to his degree of concern over the group's expectations to a significant degree.

DISCUSSION

It is to be remembered that the Ss had the criterion for a successful performance clearly in mind. They knew whether they had succeeded or failed on their individual tasks at the time they were asked to stop. With this in mind we may wonder why the Ss should pay attention to the expectations of their group.

Festinger (4) has offered a proposition which points to the salience of this question: "When an objective, nonsocial basis for evaluation of one's ability or opinion is readily available, persons will not evaluate their opinions or abilities by comparison with others." It is clear from the context of this quotation that Festinger means to include conformity to group pressures, as well as simple comparison of own performance with the achievement of others.

The results suggest that persons who were successful did not tend to evaluate their performance in the light of the wishes of the group, but that those who failed were affected by the group's expectations. Successful persons clearly evaluated their performance as highly successful regardless of whether the task was relevant or nonrelevant, the group's expectations were high or low, or the group's expectations were an important concern to them or not. Failing persons, however, evaluated their performance low if the group had high expectations, and they evaluated their performance high if the group had low expectations. The effect of the level of expectations upon the evaluation of their performance was also greater the more the persons were concerned about the expectations of the group. These results occurred when the task was relevant, but did not occur when the task was non-relevant.

Why did the wishes of the group determine self-evaluations for persons who failed more than for ones who succeeded? A failure experience, we believe, has an important property which is not present in a success. The person who has failed cannot be sure how the group will receive his report that he has failed. An audience for a report of failure, especially when it is an attractive group, is likely to create some embarrassment for the reporter. Uncertainty concerning the consequences of a report of failure makes the person who has failed highly sensitive to the opinions and evaluations of the others. Thus, social reality is more important to a person who fails than to a person who succeeds. The more the relevance of the task, the more does the person who has failed feel impelled to evaluate himself in accordance with the group's expectations. The person who has succeeded, in contrast, can be quite sure that his report of success will be

favorably received regardless of the group's level of expectations. He does not need the group for social reality and is quite ready to evaluate a success as a success regardless of the group's wishes in the matter. These results and their interpretation suggest that even though an objective, nonsocial basis for self-evaluation is available, a person may evaluate himself by comparison with others, if he has failed on a task which is relevant to the group.

It is noteworthy that relevance or non-relevance of the task, exclusive of the group's expectations, did not affect the evaluation that a person placed upon a failure experience. These results indicate, we believe, that "letting a group down" on a relevant task is no more conducive to poor self-evaluation than failure on a nonrelevant activity. The evaluation a person assigns to his performance is primarily affected by the group's level of expectations, where the task is relevant.

The results concerning the effect of self-esteem suggest several interesting issues. Why, for example, did self-esteem affect the member's evaluation of his achievement more when he had failed on a nonrelevant task than when he had failed on a relevant one? If it is assumed, as we have done, that group pressures are less potent in the nonrelevant condition, it is probable that the influence of self-esteem upon personal evaluation was relatively stronger than the power of the group in that condition. In the relevant condition, however, both because relevance increases the power of the group, and because a failure experience made the group more potent as a scale of reference, self-esteem was relatively weaker than the group as a determinant of the self-evaluation. This belief has been supported by the findings from several studies, as yet unpublished (1, 8). In these investigations, variation in the behavior of Ss, which was predicted because of variation in their personality characteristics, occurred more often when influences from the group were weak than when influences from the group were strong.

The results indicate that persons with high self-esteem, as we have conceived of it and measured it, dealt with their experiences in the laboratory in a way that helped them to maintain their high self-esteem. They tended to evaluate themselves better after a failure than

individuals with low self-esteem, regardless of the level of the group's expectations. Furthermore, they became more concerned about the group's expectations when favorable evaluation was readily possible for them. Persons with low self-esteem reacted to their experiences in a way that made it difficult to improve their self-esteem. They not only evaluated a failure very low, but they became more concerned with the group's expectations when an unfavorable evaluation was most likely for them. Persons with high self-esteem, in brief, reacted to their experiences in the group in such a way that they best protected themselves from unfavorable evaluation.

SUMMARY

In a laboratory experiment, members were made to believe that their group had either high or low expectations for the quality of their achievement on individual tasks. The tasks were described as either relevant or nonrelevant to the purposes of the group, thus varying the power of the group over the member's activity. Half of the Ss within each of these four conditions were allowed to succeed and the other half were made to fail. In addition, measures of self-esteem were available for all participants.

The following hypotheses were supported.

1. A level of aspiration set by a group for its members is more often accepted by a member as a personal goal when the activity is relevant to the group than when it is nonrelevant.
2. A failure on a relevant task generates poorer self-evaluation of a person's performance if a group's aspirations are high than if they are low (while the evaluation of a failure on a nonrelevant task is not affected by different levels of the group's aspirations).

2-a. A failure, when the person accepts the group's level of aspiration as his own, generates lower evaluation of his performance if the group's aspirations are high than if they are low (while evaluation of failure when the group's aspiration is rejected is not affected by different levels of the group's aspirations).

Hypotheses concerning the effect of the level of the group's aspirations upon the evaluation of a successful performance were not supported. It was proposed that a person is more likely to evaluate a failure than a success

in reference to the group's expectations; because the failure makes him more dependent upon the group for social reality concerning evaluation, than does success.

Persons high in self-esteem protected themselves against unfavorable evaluations better than those low in self-esteem.

The group's expectations appear to have been more potent as a scale of reference than the individual's self-esteem in determining his evaluation of his performance. When the influence of the group was weakest (task was nonrelevant) persons high in self-esteem, compared to those who were low, differed in the way they evaluated their performance. When the influence of the group was strongest (task was relevant) there was no difference in the way that persons high or low in self-esteem rated their achievement.

REFERENCES

1. COHEN, ARTHUR R. The effects of individual self-esteem and situational structure on threat oriented reactions to power. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
2. HERTZMAN, M., & FESTINGER, L. Shifts in explicit goals in a level of aspiration experiment. *J. exp. Psychol.*, 1940, **27**, 439-452.
3. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271-282.
4. FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117-140.
5. LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
6. RASMUSSEN, G., & ZANDER, A. Group membership and self-evaluation. *Hum. Relat.* 1954, **7**, 239-252.
7. SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190-207.
8. THOMAS, E. Effects of role interdependence and ego strength on group functioning. Unpublished doctor's dissertation, Univer. of Michigan, 1956.

Received February 29, 1956

AGE DIFFERENCES IN VERBAL LEARNING¹

SHELDON J. KORCHIN AND HAROLD BASOWITZ

Institute for Psychosomatic and Psychiatric Research and Training, Michael Reese Hospital

THE aphorism that "you can't teach an old dog new tricks" undoubtedly predates psychological research both in the field of aging and of learning. However, the specific conditions which determine the relative ease or difficulty of learning for the older person, the importance of different types of material, learning situations and the like, have been a concern of systematic and experimental work within the psychology of learning only in relatively recent years.

Of considerable interest is a 1934 study by F. L. Ruch (13) which has since become a basic reference in discussions of age-related changes in human learning. With the view that senility can be conceived as a condition of lowered biological plasticity, Ruch hypothesized that older compared to younger persons should show a greater deficit in any learning situation which requires the reorganization of existing habits, less deficit where earlier experiences can be utilized in new learning, and have an intermediate amount of deficit where old habits neither aid nor block new learning.

In his experiments Ruch (13) had three age groups learn a pursuit rotor task, first in direct vision, then in mirror vision, and three lists of paired-associate materials. These lists differed in the degree to which they utilized or contradicted earlier learning: (a) paired associates of familiar words, logically related; (b) nonsense materials (arbitrary equations of the form $E \times Z = G$); and (c) interference materials (false equations of the form $3 \times 5 = 6$). These tasks were presumed to differ not only in difficulty and novelty but also in the degree to which prior experience might enter to facilitate or disrupt present learning. In the word-associate task earlier learning should facilitate

the learning of these simple associates; the nonsense equations required the learning of essentially new material which should be unaffected by earlier information; while in the case of the false equations the material was not only new but contradictory to previously-formed familiar associations. With such material one might expect competing responses, the highly overlearned true products of simple multiplications, to interfere with the acquisition of the new habit.

Comparing the differences in mean scores between age groups on each of the procedures, Ruch (13) found that the young-old difference was greater for the mirror than direct vision perceptual-motor task and, among the verbal tests, was greatest for the interference material, less for the nonsense task and least for the familiar word associates. On this basis Ruch concludes that older persons have least "deficit" in learning materials which are compatible with habitual behavior and greatest where the new learning contradicts earlier habits. However, if one compares the scores for each procedure within each group separately, it is interesting to note that all groups performed *better* on the interference material than on the nonsense task, although the *difference* between young and old is greater on the former than on the latter.

As part of a larger study of psychological and neurophysiological processes in aging, Ruch's three verbal learning procedures were used in the present study. The findings of other studies in this series are reported elsewhere (2, 10). The present paper will consider only the comparison of two groups, distinct in age, in their performance on these procedures. The purpose is twofold: (a) To repeat Ruch's study and to retest his hypothesis; and (b) To see whether further understanding can be gained about age-related changes in cognitive functioning generally, and as they may account for performance in these learning tasks.

METHOD

Subjects

Two distinct age groups were used in the study. The young group consisted of 24 Ss between

¹ This study was supported by the U. S. Army under Contract No. DA-49-007-MD-261 through the Medical Research and Development Board, Office of the Surgeon General, Department of the Army, Washington, D. C. We are grateful to Mr. Ben Grossman, Director, and the staff of the Drexel Home for their complete cooperation in providing subjects and facilities. Dr. Albert Caron and Mrs. Ann Gunther were responsible for much of the data collection and Dr. Helen Heath for the statistical analyses.

the ages of 22 and 33 years of age, and included 12 resident physicians with a mean age of 28.2 years and 12 graduate nurses with a mean age of 26.3 years. The old group consisted of 36 residents of an institution for the aged, evenly divided by sex within three age groups: 12 Ss between 65-74; 12 between 75-84; and 12 above the age of 85. The Ss were included in the older sample only if they were in good general health, free from organic brain pathology, and had acceptable corrected vision.

However, the total young group was distinctly superior to the old in education and intellectual level. Since one might reasonably expect intelligence, in general, to be related to performance on these tasks and, in fact, Ruch (13) had found this to be true, particularly within his older groups, it was felt that a more direct and exact test of the differences in learning could be made using groups equated for general intelligence. Since vocabulary correlates well with general intelligence as measured by omnibus batteries and itself shows relatively little change with age, the Vocabulary test of the Wechsler-Bellevue Scale was used (1, 4, 8, 15). Two matched groups of 16 Ss each were drawn from the larger sample. The young group had a mean unweighted vocabulary score of 26.8 with a standard deviation of 4.6; the old group 28.6 and 4.8, respectively. The young group consisted of 5 men and 11 women, with a mean age of 26.8; the old group of 8 men and 8 women, with a mean age of 78.1 years.

The results of this experiment will be reported for these selected groups matched in intelligence. The same analyses have been carried out, however, for the total groups and the results are quite similar, although in the main the differences between the larger groups are more extreme.

The Learning Tasks

As part of a larger battery of procedures, each S was individually administered a modified version of the learning tasks described by Ruch (13). As in his experiment, Ss had to learn three separate lists: 1. *Word Associates*, consisting of familiar, logically related word pairs such as Man-Boy; 2. *Nonsense Equations*, e.g., $E \times Z = G$; 3. *False Equations*, e.g., $3 \times 5 = 6$. In the present study the task was changed somewhat from Ruch's original procedure. After preliminary tests, the length of list was reduced from 10 to 8 items and

the number of trials from 15 to 6 in order to make the task easier and the session less fatiguing, particularly for the older Ss.

Each list consisted of 8 units (Table 1) selected from the 10 used originally. Before each list was given S was first shown a sample pair and the task explained. After an initial presentation of the series, S was given 6 trials in which he was to anticipate the right-hand member of the pair when the left was first shown alone. The items were projected by a 35 mm. strip film projector in a semidarkened room, each frame remaining in view for 4 seconds. The left-hand member of the first pair appeared alone for 4 seconds, both parts for 4 seconds, then the left-hand member of the second pair, and so on. Eight seconds were allowed between trials and about 2 minutes between lists. To prevent serial learning, the order of items was changed on each trial.

The instructions used were substantially the same as Ruch's. The S was encouraged to guess if uncertain. In presenting the lists of Nonsense and False Equations the items were divided at the equal sign, e.g., first $3 \times 5 =$ was shown, then $3 \times 5 = 6$.

RESULTS

In Table 2 are presented the mean scores for each of the three tests. Since there were 8 items and 6 trials, the highest possible score for an individual is 48. Young Ss clearly perform better on each task. The difference between the groups is significant in each case at the $p < .001$ level on *U* test. Moreover, it is equally clear that for each group the Word-Associate learning was decidedly easier than either the learning of the Nonsense or False Equations, although, surprisingly, there is virtually no difference between the latter two.

Following Ruch's (13) analysis for determining variation in the amounts of deficit among the three types of material, we can compare the difference between old and young on each of the tests. For the Word Associates this difference in score is 13.81, for the Nonsense Equations 17.57, and for the False Equations it is 17.62 (Table 2). These values do not differ significantly from each other.

However, the difference between the groups in the learning of the three types of material

TABLE 1
LISTS USED IN THE THREE LEARNING TASKS

	Word Associates	Nonsense Equations	False Equations
1.	Man-Boy (example)	$E \times Z = G$ (example)	$3 \times 5 = 6$ (example)
2.	Stem-Bud	$R \times S = Q$	$2 \times 5 = 8$
3.	House-Visit	$H \times Y = J$	$6 \times 3 = 5$
4.	Room-Light	$B \times D = M$	$5 \times 1 = 7$
5.	Soft-Chair	$A \times M = B$	$3 \times 1 = 1$
6.	Tree-Flag	$S \times Q = H$	$3 \times 3 = 4$
7.	Nest-Owl	$K \times U = T$	$5 \times 5 = 11$
8.	White-Pink	$W \times L = S$	$2 \times 4 = 9$
	Walk-Car	$L \times B = D$	$3 \times 4 = 2$

TABLE 2
MEAN LEARNING SCORES FOR YOUNG AND OLD GROUPS

Group	N	Mean Age	Word Associates (WA)	Nonsense Equations (NE)	False Equations (FE)	WA-NE	WA-FE
Young	16	26.8	44.19	25.00	24.81	43.59	44.21
Old	16	78.1	30.38	7.43	7.19	74.86	74.96
Diff _{M_y-M_o}			13.81	17.57	17.62		

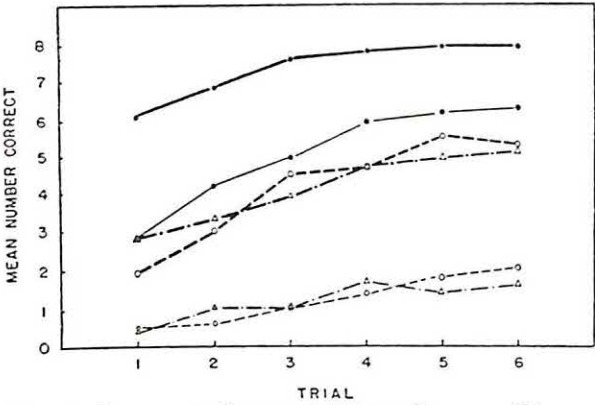


FIG. 1. LEARNING CURVES FOR THE OLD AND YOUNG GROUPS FOR EACH OF THE THREE TESTS

The curves for the young are heavily drawn, for the old lightly drawn. The three tests are distinguished: — Word Associates; - - - Nonsense Equations; - · - · - False Equations.

can be viewed in another way. If we consider the proportional change between the tests it does appear that the older group has relatively poorer learning on both the Nonsense and False Equations when these are compared with the Word Associates. Subtracting the Word-Associate score and the Nonsense Equation score and taking this difference as a fraction of the Word-Associate score, the mean ratio $\left(\frac{WA - NE}{WA}\right)$ is 43.59 for the young Ss and for the old 74.86. The difference between these values is significant, on *U* test, at the $p < .001$ level. Similarly, the ratio of Word-Associate score minus False Equation divided by Word Associate $\left(\frac{WA - FE}{WA}\right)$ is 44.21 for the young group and 74.96 for the old (Table 2); on *U* test significant at the $p < .001$ level. From this analysis we may fairly conclude that not only do the older Ss learn less in each of the three tasks, but that they are proportionately more deficient in the learning of materials in which the facilitative effects of prior experience are minimized. However, we have not found evidence that interference materials are more difficult than nonsense materials.

The trial-by-trial learning curves for the two groups on each of the tests show the difference in performance already discussed to be visible throughout learning (Fig. 1). The curves for the Nonsense Equations and False Equations are virtually identical within each group and each of these is distinct from Word-Associates levels. It does seem, however, that the older

group's curves for the Nonsense and False Equation tasks approximate linear functions while their values on the Word Associates more nearly describe the more usual negatively accelerated learning function. Particularly in the early trials there is proportionately less rise in score on the meaningless and interference materials. Using Edwards' (5) suggested analysis of variance for trend; the curves for both of these tasks differ significantly from that for the Word Associates ($p < .001$ in each case). This difference in the forms of the learning curves is suggestive of Lewis and Shepard's (11) finding of a negatively accelerated function for perceptual motor learning consistent with past experience and of a linear or positively accelerated curve in a task which did not utilize old habits.

Not only do the old and young groups differ in the amount of material learned, but they differ as well in the type of error made during the learning process. In Table 3 the total number of errors made by each of the groups on each of the three tasks is broken down into various types. In both the Word Associates and the Nonsense Equations there are three distinct types of error possible: *S* may not respond at all to the stimulus item (*No response*); he may give an associate which appears nowhere in the series (*Random response*); or he may respond with an item which although incorrect for the present stimulus item is appropriate elsewhere in the list (*Correct elsewhere*). For the False Equations list *S* can also give a response which represents a correct multiplication, or conceivably, he can add or subtract the two numbers in the stimulus equation. Although the completion of the multiplication is of particular theoretical interest since it may repre-

TABLE 3
PERCENTAGES OF VARIOUS TYPES OF ERRORS MADE IN WORD ASSOCIATE (WA), NONSENSE EQUATION (NE) AND FALSE EQUATION (FE) TESTS

Group	N	Mean Age	Test	Total No. of Errors	No Response	Random Response	Correct Elsewhere	Multiplication*	Addition	Subtraction
Young	16	26.8	WA	61	62	7	31			
			NE	368	45	7	48			
			FE	370	30	9	45	2	8	6
Old	16	78.1	WA	273	83	1	16			
			NE	649	92	1	7			
			FE	653	75	3	18	1	1	2

* These types of errors are only possible on the False Equation task.

sent the intrusion of the competing, albeit actually correct, response, addition or subtraction may also reflect habitual response to the implied structure of an arithmetical problem. Scoring these errors does, of course, involve interpretation since such response might as well be a *Random response* or, in a particular case, *Correct elsewhere*.

There is considerable difference in the pattern of errors between the two groups (Table 3). On all three tasks the typical behavior of the older person is not to respond, distinctly less often to give a response that is *Correct elsewhere*. Compared to the old group, young Ss have proportionally more of their errors of the latter type. They are also more prone to give *Random responses* and in the case of the False Equations to multiply, add or subtract the items, although there are few of these errors in either group.

Closer inspection of the trial-by-trial changes in the pattern of error gives further evidence of the difference between old and young learners. On all tasks on each successive trial the young group has relatively fewer of its errors in the *No response* class and an increasing proportion of *Correct elsewhere* errors. On the Nonsense Equations task, for example, of the total errors made by the young Ss on the first trial 57% are *No response* and 30% *Correct elsewhere*, which changes gradually to 30% *No response* and 66% *Correct elsewhere* by the sixth trial. However, the old group preserves roughly the same proportions on each trial. Thus, for the same task, these Ss start with 95% *No response* and 3% *Correct elsewhere* and by the last trial still have 92% and 8% of their errors in each of these classes. Although there is somewhat more tendency for the *No response* proportion to go down on the Word Associate and False Equation tasks, the pattern is basically the same. Thus, it is clear that simply not responding to a stimulus item is characteristic of the learning behavior of the older Ss throughout; it is less common in the younger Ss at any point and declines greatly as learning proceeds.

Lastly, how related are the performances of individuals on the three learning tests? The correlations among the total scores for the three tests within each of the age groups are given in Table 4. The results are surprising. For the young Ss there is a positive relationship among all the procedures, lowest between the Word

TABLE 4
CORRELATIONS AMONG LEARNING TASKS

	Young	Old
Word Associates vs. Nonsense Equations	.372	.506*
Word Associates vs. False Equations	.607*	.317
Nonsense Equations vs. False Equations	.740**	-.240

* Significant at $p < .05$.

** Significant at $p < .001$.

Associates and Nonsense Equations, and quite high between the two forms of equations. However, the older group has its highest correlation between the Word Associates and Nonsense Equations and its lowest, and in a negative direction, between the Nonsense and False Equations. Thus, the patterns are opposite in the two groups. Moreover, these findings differ sharply from those of Ruch who found that all of the interest correlations were higher in his oldest group and highest among these was the correlation between the Nonsense and False Equations. Indeed, the relationship between these two tests was generally stronger than between any other pair in his experiments: +.404 in the young group; +.473 in the middle-aged; and +.833 in the old. We are at a loss to explain the differences in findings both between our two groups and between this experiment and Ruch's.

DISCUSSION

There is no doubt that older persons in this as in Ruch's (13) original study are less well able to master each of the learning tasks. As such, this finding is in agreement with a considerable body of material which shows a generally decreased ability to learn with advancing age (6, 12, 14). However, unlike Ruch, we have found virtually no difference between scores on the nonsense and interference materials and no evidence of a greater degree of deficit in the latter than in the former. As was noted earlier, the total amount learned by Ruch's Ss of all ages was greatest on the Word Associate task, second most on the False Equations and least on the Nonsense Equations, while deficit, i.e., difference in performance of young and old, was least for the familiar material, greater for the nonsense material, and greatest for the interference material. However, it may be that the differences, though seem-

ingly small, between Ruch's and the present procedure are sufficient to account for the obtained differences in results. Perhaps with longer lists and more trials the early findings would have been repeated, but in any case the differences in method do limit direct comparison of results.

What we have found is that, compared to the level of attainment on Word-Associate learning, the older persons show proportionately poorer learning on *both* the new and interference materials. In this sense, we can speak of greater deficit on both of these materials, although there is no difference between them. While it may be that older persons are uniquely prone to disturbed performance when required to reorganize existing habits, as has been suggested by a number of papers (7, 9, 14, 16), the evidence of the present experiment suggests that the older individual will perform less well in any situation in which he must deal with material which is in any way novel. When the task at hand can not be integrated with earlier experience, either because there is no precedent or existing precedent is inadequate or contradictory, learning suffers. The lessened capacity of the aged to deal with novelty was also found in heightened decision times at the beginning of a series when *S* had to judge a number of ambiguous stimuli (10).

Older and younger *Ss* differ also in their behavior during learning. The younger, as we have noted, are more likely to make some response to the stimulus item, to offer hypotheses, whether correct or incorrect, and as learning proceeds to have an increasing proportion of their errors responses which, although correct elsewhere in the list, are inappropriately given to the particular stimulus. The older *S* will either respond correctly or not at all, and this behavior is consistent from trial to trial. In a study of the perception of closure the older *Ss* also showed a greater tendency not to respond to Gestalt Completion Test items (2).

We believe that three factors jointly determine this behavior of the older *Ss*, although from available evidence we cannot judge their relative contributions.

1. The older person is more cautious. Despite the injunction to guess, it may be that older persons require greater certainty before they are willing to report. Although *S* knows that the item is equally incorrect whether he gives

no response or a wrong response, he apparently prefers the error of omission to that of commission. This tendency to inhibit response in the uncertain situation may reflect a more profound personality defense of the aged through which the recognition of inadequacy is avoided.

2. The older person requires more time for the integration of a response. A condition of the experiment which may account for some of the differences may be the rapid pacing of the items. As the experiment was conducted *S* had to read the stimulus item, select and report a response item within 4 seconds. Considerable work has shown that older persons require more time both for simple perceptual discrimination and in more complex decision-making (3). Moreover, requiring performance within arbitrarily short times can result in a significant reduction in the adequacy of that performance (16). Such findings would suggest that the processes of perceiving-judging-recalling-responding require more time for the older person than is available and that, perhaps, if the pacing were slower his behavior might differ less than that of younger individuals.

3. For the older person learning may principally involve the association of *discrete* stimulus-response combinations, and unlike the younger person, may not include a prior stage in which the response items are first learned as belonging *somewhere* in the series, before they are attached to particular stimuli. That the pattern of errors is generally the same on all three learning tasks and that there are no important changes in pattern from trial to trial further suggest that the acquisition of separate and distinct units may be a general characteristic of learning in the aged. Implicit in this view is the assumption that there are in general two patterns of learning: one in which there is first a more general learning from which particular associations are increasingly differentiated; and a second in which only the specific stimulus-response associations are formed. We might speculate that the second pattern, seen here in the behavior of aged persons, is generally characteristic of more constricted or rigid cognitive organization and might be found as well in the behavior of anxious or brain-damaged individuals.

From the first of these three views one assumes that the older learner may have available, but is unwilling to report, incorrect hypotheses. According to the second, the

formulation of such hypotheses can not be completed and reported within the time available. But on the last assumption neither caution nor time blocks response, rather there are fewer incorrect hypotheses available. The relative importance of these factors could be explored in further experiments in which, for example, Ss might be required to make some response to each item and/or the time of presentation increased.

All of these factors, however, may be conceived as parts of the total aging process. In the aged there is a decreased ability to comprehend, organize, and integrate new experience both in perception and in memory. Integration is likely to be incomplete, often rigid or confused, and the integrative process requires greater time. Performance is most adequate where the individual can draw most heavily from past experience, most disturbed when novel behavior is involved. As a consequence the aged person seeks the security of the familiar, risks fewer failures by making fewer attempts, and so defends against the recognition of inadequacy.

SUMMARY

An old and a young group of Ss, mean ages 78.1 and 26.8 years, were given three paired-associate learning tasks which differed in the degree to which prior experience might be expected to facilitate or block present learning. They consisted of (a) familiar word pairs, (b) nonsense equations, and (c) false equations.

Both groups performed best on the word-associate task, but there was little difference between the learning of nonsense and false equations within either group. On all three procedures the old group was significantly poorer, but they were proportionately more deficient in the learning of materials in which the facilitative effects of prior experience are minimized, i.e., the two forms of equations. However, they had no greater difficulty with the interference than with the nonsense material. The older person thus seems to be less capable in dealing with novel material which can not be readily intergrated with earlier experience.

Older and younger Ss were also found to differ in their behavior during learning. Younger Ss are more likely to offer hypotheses, whether correct or incorrect, and through learning to have an increasing proportion of

their errors responses which though correct elsewhere in the series are incorrect for the particular stimulus. By contrast, the old Ss respond correctly or not at all, and the proportion of "no response" errors remains high through learning. Three factors are discussed as joint determiners of this behavior: older persons are more cautious; they require more time to integrate response; and their learning may consist mainly of discrete stimulus-response associations, not preceded by more general learning. These factors are conceived as related parts of the total aging process.

REFERENCES

1. BABCOCK, HARRIET, & LEVY, L. *Test and manual of directions, the revised examination for the measurement of the efficiency of mental functioning*. Chicago: C. H. Stoelting, 1950.
2. BASOWITZ, H., & KORCHIN, S. J. Age differences in the perception of closure. *J. abnorm. soc. Psychol.*, 1956, **54**, 93-97.
3. BIRREN, J. E. Age changes in speed of simple responses and perception and their significance for complex behavior. In *Old age and the modern world*. Report of the Third Congress of the International Association of Gerontology. London: Livingstone, 1955. Pp. 235-247.
4. BOTWINICK, J., & BIRREN, J. E. Differential decline in the Wechsler-Bellevue subtests in the senile psychoses. *J. Geront.*, 1951, **6**, 365-368.
5. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
6. GILBERT, JEANNE G. *Understanding old age*. New York: Ronald, 1952.
7. HALSTEAD, W. A psychometric study of senility. *J. ment. Sci.*, 1943, **89**, 363-373.
8. HOWELL, R. J. Changes in Wechsler subtest scores with age. *J. consult. Psychol.*, 1955, **19**, 47-50.
9. KAY, H. Learning of a serial task by different age groups. *Quart. J. exp. Psychol.*, 1951, **3**, 166-183.
10. KORCHIN, S. J., & BASOWITZ, H. The judgment of ambiguous stimuli as an index of cognitive functioning in aging. *J. Pers.*, 1956, **25**, 81-95.
11. LEWIS, D., & SHEPARD, A. H. Prior learning as a factor in shaping performance curves. *Proc. nat. Acad. Sci.*, 1951, **37**, 124-131.
12. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
13. RUCH, F. L. The differentiative effects of age upon human learning. *J. gen. Psychol.*, 1934, **11**, 261-268.
14. THORNDIKE, E. L., BREGMAN, ELSIE O., TILTON, J. W., & WOODYARD, ELLA. *Adult learning*. New York: Macmillan, 1928.
15. WECHSLER, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
16. WELFORD, A. T. *Skill and age*. London: Oxford Univer. Press, 1951.

Received March 1, 1956

PERSONALITY, PROBLEM SOLVING, AND THE EINSTELLUNG EFFECT¹

BRENDAN A. MAHER

Northwestern University

ADORNO (1) and others have reported a pattern of personality characterized by certain cognitive and behavioral tendencies among which is included rigidity of set and outlook (1, p. 561). The concept of rigidity has been employed in many ways, a comprehensive summary of which is to be found in a paper by Fisher (6), but a review of its uses is unnecessary here. This study is concerned with rigidity as defined by Adorno and his colleagues, ("pre-conceived categorizing, inaccessible to new experience") in relation to certain other aspects of personality, in particular pseudo-scientific or antisocial tendencies in interpreting individual behavior.

The "ethnocentric" individual is described by the same writers as exhibiting a cognitive personality organization that includes "implicit denial of personality dynamics, ready explanation by accidental factors, heredity, etc., superstition." This method of organizing perceptions of the self and others is defined as antisocial and is contrasted with the tendency to explain human behavior in dynamic sociopsychological terms.

Rokeach (12) has used the familiar water-jar problems to demonstrate that rigidity of set in problem solving is found more frequently in "high" ethnocentric groups than in "low" groups. His findings have been challenged by Luchins (11) and Brown (2); the later concluding that this relationship existed only "... in ethnocentric Ss when they felt an anxious concern over achievement."

Against this background the present study was designed to test two simple hypotheses: (a) Ss who reveal antisocial or nondynamic tendencies in their perceptions of human behavior are more subject to the Einstellung effect in cognitive problem solving than Ss who use dynamic psychological concepts in interpreting human behavior. (b) This relationship

is more evident when the Ss are placed in a situation likely to give rise to achievement anxiety.

The use of "antisocial" bases for evaluating individual behavior has been described in an unpublished study by Sullivan (13). He used a method in which Ss were required to sort and group photographs of people on any basis that they might wish to use. The sorting categories reported by the Ss were then classified as objective or nondynamic concepts (e.g. hair-color, physical characteristics in general, etc.), or as inferential (dynamic) concepts, the latter implying some attempt to use inferred psychological or behavioral characteristics. His findings were the reverse of those to be predicted on the basis of the previous studies in ethnocentrism. It was suggested post facto that in this experiment the "scientific" or objective approach would properly utilize nondynamic categories, as the judgment of personality from photographs is essentially an unscientific procedure based on a willingness to infer personality attributes on the basis of pictured, external characteristics.

To avoid this ambiguity of interpretation, the present study used a measure derived from Kelly's Role Construct Repertory Test (RCRT) (9). This procedure presents S with the task of evaluating particular individuals with whom he is personally acquainted, and may therefore be assumed to elicit the perceptual categories that he uses in everyday life. In the test, S lists a number of persons of his acquaintance, on the basis of definitions (Kelly's "role-titles") provided for him. He is then required to consider particular trios of these individuals in turn. Each trio is divided by S into a single and a pair of individuals on the basis of some perceived similarity between the two members of the pair which is not shared by the third (single) member. For each trio, S reports the basis upon which the division was made. The list of definitions is the same for each S, but the persons actually listed will, of course, be different. For any particular purpose the number and nature of the definitions may

¹ The data upon which this paper is based are taken from a dissertation submitted by the writer to the Graduate School of the Ohio State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Particular acknowledgments are due to Dr. George Kelly for his advice and encouragement.

be altered, and the number of trios may be of any size within the limits of combinations permitted by the number of definitions.

One estimate of the reliability of the RCRT has been derived by Hunt (8). He administered the test to a sample of college students followed one week later by a similar administration but using different role titles. A mean percentage agreement of 70.1 was obtained. As responses were only regarded as in agreement when they were identical, it is apparent that any broader definition of similarity could not reduce Hunt's percentage and would be likely to increase it.

METHOD

Subjects and Procedure

Undergraduates in an elementary mental hygiene class were divided into two equal groups.

Achievement anxiety group. These were students in one section of the course. At the beginning of the course they completed the RCRT. Five weeks later a mid-quarter examination was held in which all Ss participated as a course requirement. At the next class meeting the group was informed that the examination results had been very unsatisfactory and that as an aid to grading them two simple tests of "mental alertness" would be administered immediately. The problem-solving tests were then administered in group fashion. At the end of the class hour their graded examination papers were returned to the group and the experiment explained. This procedure was intended to demonstrate that the problem-solving tests had not actually affected their examination grade and to eliminate any residual anxieties. It was E's impression that during the test session the motivation of the group was sufficient to justify the inference that achievement anxiety was a relevant variable.

Control group. This group, which comprised another section of the same course, completed the RCRT as above. The problem-solving tasks were administered in group fashion during a regular class meeting. They were presented with standard instructions as part of the demonstrations used regularly in the course.

This procedure was repeated over two academic quarters because of the attrition rate associated with errors in the completion of the printed problems. Complete data were obtained from 62 Ss, 31 in each group.

Instruments

Verbal maze problems. This series of problems was identical in content and sequence with that reported by Cowen² (5). Each maze was ditto-printed on a separate page. The pages were stapled together in such a fashion that it was necessary to solve each maze in turn and then detach it before being able to inspect the next problem in the series. Following Cowen's method, the problems were presented in sequence commencing

TABLE 1
EINSTELLUNG PROBLEMS: WATER-JARS SERIES

Problem Number	Sizes of Jars Available		Quantity to be Obtained	Type of Problem
A	39	4	31	2 jar practice
B	30	40	3	3 jar practice
1	27	74	20	Control
2	42	71	5	1st Einstellung
3	50	81	7	2nd Einstellung
4	23	57	9	3rd Einstellung
5	27	71	14	4th Einstellung
6	35	77	17	5th Einstellung
7	21	51	9	1st Crucial
8	32	79	15	2nd Crucial
9	37	90	16	3rd Crucial
10	31	75	13	4th Crucial
11	19	65	7	Extinction
12	28	65	9	5th Crucial

with two practice problems; one control problem (which could be solved by a long or a short method of solution); five Einstellung problems soluble only by the long method; four crucial problems which, like the control problem, were amenable to either method of solution; one extinction problem, not soluble by the long method; and finally one more crucial problem. The Einstellung score is the number of crucial problems solved by the long method, and could attain a maximum value of 5.

Water-jar problems. These problems were presented in the same control-crucial sequence as the verbal mazes. The problems used are given in Table 1 below. Administration of the problems was in booklet form to the groups of Ss in a manner identical with that used in the verbal mazes. To take into account the possibility that perception of the correct solution in one series of problems may transfer to the solving of the second series, one half of each group commenced with the verbal mazes and then went on to the water-jars problems. The sequence was reversed for the other half of each group. Thus while transfer effect was not eliminated, it is presumed to have influenced the group mean scores equally for both problems.

Role Construct Repertory Test. The RCRT was administered with 11 definitions and 22 triads. The definitions used were as follows:

1. "Myself"
2. Father or surrogate
3. Mother
4. Brother
5. Sister
6. Opposite sex peer most understood
7. Opposite sex peer least understood
8. College professor most understood
9. College professor least understood
10. Same sex peer most understood
11. Same sex peer least understood

The triads were selected randomly with the restriction that each defined figure was to be categorized six times.

From each S, a total of 22 category responses were obtained, with the possibility that any category might appear more than once. The categories used by each S were typed on separate cards and numbered to identify the protocol from which they were taken. As a

² The complete set of mazes was supplied to the writer by Cowen in a personal communication.

result of the considerable number of duplications found within and between protocols, the total of categories was well below the possible maximum of 1,364. The cards were then inspected and sorted into two groups by each of 3 judges in turn, on the basis of the following criteria.

Nondynamic categories. A group including all those categories based upon: (a) Physical appearance, health, height, weight, etc.; (b) Kindred relationship, degree of acquaintance, etc.; (c) Legal status, sex, age, occupation; (d) Geographical, cultural, or social group membership; (e) Presence or absence of a specific skill or achievement.

Dynamic categories. A group including all categories referring to personality in terms of needs, motives, drives, characteristic behavior patterns, etc. Examples would be "aggressive," "intelligent," "ambitious," "anxious" and the like.

The judging was on a forced-choice basis, as each category card was to be placed in one of the two groups. Where differences arose between the judges' evaluations, the majority decision was accepted. Each judge operated independently, the collating of judgments being done by the experimenter. Interjudge agreement was computed in terms of percentage of agreement on the total cards judged. The lowest agreement between any two judges recorded was 91.1; the highest was 99.0. Under the circumstances it seems reasonable to conclude that the criteria are reliable as bases for judgment.

The judging completed, each *S* was then scored with respect to the number of times which he used nondynamic categories to discriminate between figures. The maximum possible score was 22. Scores obtained ranged from 0 to 19. A range of 0 to 3 included the 31 lowest scores, while the high scores ranged from 5-19. The mean scores of these two groups were 1.8 and 9.3 respectively.

RESULTS

The mean Einstellung scores of both groups on the two problem-series are presented in Table 2. Analysis of these data revealed a statistically significant difference in respect to

TABLE 2
MEAN EINSTELLUNG SCORES IN RELATION TO
ACHIEVEMENT ANXIETY AND TO USE OF
NONDYNAMIC (ND) CATEGORIES

Groups	N	Einstellung Test Means	
		Water jars*	Mazes**
Achievement anxiety			
High ND	17	2.13	2.20
Low ND	14	2.25	2.07
Control			
High ND	14	1.55	1.80
Low ND	17	1.44	1.63

* For achievement anxiety vs. control, $t = 2.29, p < .05$; for High ND, $t = 1.99$ (n.s.)

** For achievement anxiety vs. control, $t = 1.97$ (n.s.); for High ND vs. Low ND, $t = 1.3$ (n.s.)

the achievement-anxiety variable on the water-jars problems. There were no other significant differences between means. In no instance did an *S* use the long method for solution of the final (postextinction) crucial problem.

In view of the fact that the High Nondynamic (ND) grouping includes a wide range of scores, it was deemed advisable to ensure that the differences in problem scores were not being contaminated by initial differences in ND scores between the Control High ND and the Achievement-Anxiety High ND subgroups. Therefore the 20 highest and 20 lowest ND scores were selected and their distribution between the four groups was tabulated. A chi square derived from this distribution was nonsignificant. It is therefore assumed that the differences found between groups can be attributed to the obvious variables and not to chance maldistribution of the High ND *Ss* between the control and experimental conditions.

DISCUSSION

The data obtained in this study do not support the hypothesis that individuals who show nondynamic tendencies in their perceptions of other people are more susceptible to the Einstellung effect in cognitive problem-solving situations. Some support is offered for the conclusion that the presence of achievement anxiety creates this susceptibility in *Ss* irrespective of their tendencies to dynamic or nondynamic perception in interpersonal relations. No evidence of any interaction effects was found and the obtained difference was significant only for the water-jars problems.

These findings appear to be relevant to the broader issue of generality in rigidity. The issue is, essentially, whether rigidity is to be regarded as a dimension of a given person's behavior that underlies many diverse actions, from social relationships to mathematical calculations, or as relatively specific to a given situation. The latter alternative implies that rigid behavior might be elicited from anyone in that situation. It appears that the data reported above support the latter rather than the former position. In this respect they run counter to the findings of Brown in a similar experimental situation (2). Brown reports a significant correlation between scores on the *F* scale and Einstellung scores when *Ss* were

working under conditions designed to arouse achievement anxiety, but not under "relaxed" conditions, with the implication that rigidity functions as a "preferred" defense for some persons, being elicited by any stimulus perceived as threatening to status or prestige. In the absence of such threat, the rigidity would be either minimal or absent.

The fact that this relationship was not found in the present study may be due to a number of factors. First, Brown did not use a control problem in his water-jars series as he found that its omission maximized the Einstellung effect. The fact that a control problem was used in the present study may contribute to an explanation of the different results. Whether or not to use a control problem depends upon the necessity of observing the susceptibility of Ss to set induced within the confines of the laboratory situation. A second factor is related to the arousal of achievement anxiety. In Brown's study, "the experimenter (E) maintained an extremely grave, aloof manner, dressed formally and conservatively and repeatedly cautioned Ss against looking at their test materials ahead of time," and suggested that "the tests to be taken were measures of intelligence and motivation and the results were of great importance." The present experiment, in contrast, related performance on the tests to a rather specific punishment-reward situation—namely a grade on a course in which Ss were enrolled, and may have created a higher level of achievement anxiety. Perhaps when the anxiety is sufficiently intense, even the most flexible of individuals may show rigid behaviors. When this happens, differences between individuals which may be evident in less threatening situations become obscured. Brown comments that "when the authoritarian is presented with the arithmetic problems in a very relaxed social situation he solves them as easily as the equalitarian." To this might be added the comment that when the equalitarian is presented with the arithmetic problems in a situation that is sufficiently anxiety arousing, he solves them no more easily than the authoritarian. The relationship between stress, anxiety, and rigidity has already been reported in the literature (3, 4, 7, 10).

The problem of generality in rigidity may be considered in terms of the dimensions along which rigidity is assumed to generalize. If

situations involving two or more pencil-and-paper cognitive problems show significant correlations within a given group of Ss then, it is possible to infer a factor of rigidity operating in Ss' responses to this type of problem. It may be possible to predict that rigidity will be elicited in a given S's response to other similar paper-and-pencil cognitive problems. Such evidence seems to be an inadequate basis, however, for predicting behavior in social situations. This position resembles that of Cowen, Wiener, and Hess (5) who obtained significant correlations between the water-jars problems and alphabet-mazes, but point out that such correlations cannot be regarded as evidence that there is a "relationship between rigidity and personality" except in the precise experimental situation in which the correlations were obtained. With this point of view the present writer finds himself in close agreement.

SUMMARY

Two groups of Ss, 31 in each group, completed two Einstellung-test problem series. One group was tested under standard conditions, the other under conditions calculated to arouse achievement anxiety. The Ss were then scored on a variable of dynamic or nondynamic perceptions of other people as defined by Adorno and his co-workers (1). The measure was derived from a modification of Kelly's Role Construct Repertory Test. This variable was found to be unrelated to the Einstellung-test scores, whereas a significant relationship was found to achievement anxiety. The results are interpreted as generally favorable to the view that situational or field variables play a significant role in the production of rigid behavior.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.* 1953, **48**, 469-476.
3. CHRISTIE, J. R. The effects of frustration on rigidity in problem solution. Unpublished doctor's dissertation, Univer. of California, 1949.
4. COWEN, E. L. The influence of varying degrees of psychological stress on problem-solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512-519.
5. COWEN, E. L., WIENER, M., & HESS, JUDITH. Generalization of problem-solving rigidity. *J. consult. Psychol.*, 1953, **17**, 100-176.

- [6. FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
7. HARRIS, R. A. The effects of stress on rigidity of mental sets in problem solution. Paper read at A.P.A., Washington, D. C., Sept., 1952.
8. HUNT, D. E. Studies in the role concept repertory: conceptual consistency. Unpublished master's thesis, Ohio State Univer., 1951.
9. KELLY, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
10. KENDIG, ISABELLE, V. Studies in perseveration: II. Determining factors in the development of compulsive activity. *J. Psychol.*, 1937, **3**, 231-246.
11. LUCHINS, A. On recent usage of the Einstellung-effect as a measure of rigidity. *J. consult. Psychol.*, 1951, **15**, 89-94.
12. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259-278.
13. SULLIVAN, P. Conceptual properties governing the categorization of people. Unpublished doctor's dissertation, Univer. of California, 1950.

Received March 5, 1956.

ON THE RELATION BETWEEN THE PROBABILITY OF A WORD AS AN ASSOCIATION AND IN GENERAL LINGUISTIC USAGE

DAVIS HOWES

Center for Advanced Study in the Behavioral Sciences¹

THE word-association experiment, developed primarily as a clinical tool, is now being used extensively as a method for the quantitative analysis of linguistic behavior (1, 3, 5, 6). Its value in such investigations depends to a considerable degree upon the extent to which results of the experiment can be generalized to language emitted under other, more natural conditions. The criticism is often advanced, for example, that the specialized nature of the instructional set and other conditions attendant upon the experimental situation invalidates the association experiment as a technique for studying general linguistic usage.

The question can be resolved empirically only by comparing specific properties of language produced in the association experiment with corresponding properties of language produced under conditions free of the constraints of the experimental situation. One property that has been studied is the distribution of word-association frequencies, which Skinner (7) has shown to approximate Zipf's law as closely as do the distributions of word frequencies taken from samples of continuous discourse (11). The comparison is here extended from the form of the distributions to the actual frequencies with which individual words occur in both types of sample.

To perform this comparison we need a listing of all different words with their frequencies of occurrence in general usage in one column and their over-all frequencies of occurrence as responses in the word-association experiment in a second parallel column. If the correlation between the frequencies in the two columns is perfect, the language of the word-association experiment may be said to be *stochastically equivalent* to the language of general usage. In that event we could conclude that the special conditions introduced by the association experiment do not affect a person's selection of words. The effect of a stimulus word in the association experiment,

on this hypothesis, is merely to redistribute the factors underlying the emission of words without changing them. If, on the other hand, the frequencies in the two parallel columns are so disparate that it is impossible to predict one set of frequencies from the other, we would be forced to conclude that the experimental situation introduces fundamental changes into the processes underlying a person's selection of words. An empirical determination of the correlation between the two columns will permit us to decide which of these two conceptions of the association experiment applies.

DEFINITIONS

An explicit statement of the problem requires more careful definition of concepts than has been traditional in this field. In this section the concept of stochastic equivalence of language samples is developed from definitions based upon certain idealized linguistic experiments.

We restrict ourselves to the written language produced by some specified population of persons O . It is assumed that the written language emitted by O can be decomposed into sequences of occurrences of a set of discrete spatial patterns or *letters*. Any permutation of letters bounded by spaces is a *word*, and the set of all possible words for a given set of letters is the *vocabulary*. Thus for English the vocabulary is the denumerable set of words $A, B, \dots, Z, AA, AB, \dots, ZZ, AAA, AAB, \dots, ZZZ, \dots$.

A word is said to be emitted in the *free situation* if at the time of its emission the behavior of O is in no way affected by the experimenter. An occurrence of the j th word of the vocabulary is written w_j if it is emitted in the free situation.

Consider a fixed time interval T , called the *base interval*, during which a random sample from the population O emits words in the free situation. The probability of obtaining an occurrence of the j th word in a random draw from the sample of words emitted in T is the *base probability* of that word, $P(w_j)$. The base

¹ On leave from Operational Applications Laboratory, AF Cambridge Research Center. Present address: Department of Economics and Social Science, M.I.T.

probabilities of all words in the vocabulary define a vector $\mathbf{P}$,

$$\mathbf{P} = [P(w_1), P(w_2), \dots, P(w_j), \dots], \quad [1]$$

$$\sum_j P(w_j) = 1,$$

which is called the *base probability vector*. This vector is fixed uniquely by a population O and a base interval T .

Similar definitions apply to the written language produced in a word-association experiment. Let another random sample from the same population O be given a standard association experiment (4, p. 38) during randomly-selected segments of the interval T . The symbol r_j will be used to represent an occurrence of the j th word as a response (association) in such an experiment in order to distinguish it from a word emitted in the free situation. A stimulus presentation of the i th word by the experimenter is denoted by s_i . The probability of obtaining an occurrence of the j th word in a random draw from the sample of associations given in response to a stimulus presentation of the i th word is the *associative probability* of the j th following the i th word, $p_{s_i}(r_j)$. The associative probabilities of all words in the vocabulary following presentation of the i th word define a vector $\mathbf{p}_i$, called the *associative probability vector* for the i th word,

$$\mathbf{p}_i = [p_{s_i}(r_1), p_{s_i}(r_2), \dots, p_{s_i}(r_j), \dots],$$

$$p_{s_i}(r_i) = 0, \quad [2]$$

$$\sum_j p_{s_i}(r_j) = 1.$$

The i th term of this vector, $p_{s_i}(r_i)$, is always zero, since the standard experimental procedure does not permit the subject to give the stimulus word as his response.

To establish the connection between the associative vector $\mathbf{p}_i$ and the base vector $\mathbf{P}$ we consider the vector of conditional probabilities of words emitted in the free situation,

$$\mathbf{P}_i = [P_{w_i}(w_1), P_{w_i}(w_2), \dots, P_{w_i}(w_j), \dots], \quad [3]$$

$$\sum_j P_{w_i}(w_j) = 1.$$

Since

$$P(w_j) = \sum_i P(w_i) P_{w_i}(w_j), \quad [4]$$

the base vector $\mathbf{P}$ represents the weighted average of the vectors $\mathbf{P}_i$,

$$\mathbf{P} = \bar{\mathbf{P}} = \sum_i P(w_i) \mathbf{P}_i. \quad [5]$$

In order to obtain a comparable quantity for words emitted in the association experiment we define a *base associative probability vector* as the identically weighted average of the vectors $\mathbf{p}_i$,

$$\mathbf{p} = \bar{\mathbf{p}} = \sum_i P(w_i) \mathbf{p}_i. \quad [6]$$

One can think of the base associative probabilities as the frequencies of occurrence of words as responses in an association experiment in which the stimulus words represent a random sample of the words the experimental subjects would have emitted during the same interval of time had they been left in the free situation. In practice, no word is presented as a stimulus more than once in a given association experiment. Each associative probability vector $\mathbf{p}_i$ therefore has to be estimated from responses made to a single experimental presentation of the i th word. It is therefore convenient to compute the base associative vector from the product of the matrix of associative probabilities $[\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_i, \dots]$ post-multiplied by the column vector $\mathbf{P}$

$$\mathbf{p} = \bar{\mathbf{p}} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_i, \dots] \mathbf{P}. \quad [7]$$

The concepts introduced at the outset of the paper can now be given more precise definition. *Stochastic equivalence* between the language emitted in the association experiment and the language emitted in the free situation is defined by equality of the vectors $\mathbf{P}$ and $\mathbf{p}$. When these vectors are defined as in Equations 5 and 6 the equivalence is said to be of the *first order*. First-order stochastic equivalence thus implies that, for every pair (i, j) of words in the vocabulary,

$$\sum_i P(w_i) P_{w_i}(w_j) = \sum_i P(w_i) p_{s_i}(r_j). \quad [8]$$

Stochastic equivalence of any higher order can be defined by substituting conditional and associative probabilities of a higher order for those of the first order in Equations 2 and 3. Stochastic equivalence of the second order, for example, would imply in analogy to Equation 8 that

$$\sum_{ij} P(w_{ij}) P_{w_{ij}}(w_k) = \sum_{ij} P(w_{ij}) p_{s_{ij}}(r_k) \quad [9]$$

for all triples (i, j, k) of words in the vocabulary. In the present study we shall be concerned exclusively with first-order relationships. It is important, however, to recognize the possibility that the results obtained for the first order may not apply to higher-orders.

The equation $P = p$ is trivial if the associative probabilities of all word pairs are equal to their corresponding conditional probabilities, i.e., if for all pairs (i, j) of words,

$$\begin{aligned} p_{si}(r_j) &= P_{wi}(w_j), \\ p_{sk}(r_k) &= 0, \quad \sum_j p_{si}(r_j) = 1, \\ P_{wk}(w_k) &= 0, \quad \sum_j P_{wi}(w_j) = 1. \end{aligned} \quad [10]$$

Now, there are no data from which the conditional probabilities $P_{wi}(w_j)$ can be estimated accurately. But inspection of the data from any word-association experiment is alone sufficient to show that in fact associative probabilities differ greatly from conditional probabilities for at least some word pairs. To cite one example, the associative probability of the word *light* following the stimulus word *dark* is found experimentally to be about 0.4, while the probability that *light* follows *dark* in the free situation (i.e., the probability of the sequence *dark light* over the probability of *dark*) is clearly of a lower order of magnitude. Thus the real question at issue is whether the average associative probabilities are equal to the average conditional probabilities even though the individual associative and conditional probabilities differ greatly.

The method used here to test for stochastic equivalence is to calculate the coefficient of correlation between empirical estimates of $P(w_j)$ and $\sum_i P(w_i)p_{si}(r_j)$ for as large a sample of words as possible. A correlation coefficient of one implies perfect stochastic equivalence; a correlation of zero, perfect stochastic independence. The square of the correlation coefficient measures the proportion of the variance of the empirical data that is due to the relationship between the two sets of frequencies. It is thus a measure of the extent to which the hypothesis of stochastic equivalence accounts for those data. The square of the correlation also indicates the precision with which the probabilities of words in the free situation can be estimated from the laboratory situation of the association experiment. In this sense, it may be interpreted as an index of the potential value of the association experiment as a general laboratory method for the statistical analysis of language.

MEASUREMENTS

Measurements of base probabilities and associative probabilities for the same popula-

tion of subjects have not been published. For base probabilities the most extensive set of data published is the Lorge magazine count (8). This is an analysis of 4.6 million words of text from five popular American magazines issued during the period 1927-1938. For associative probabilities the standard tables are still those published by Kent and Rosanoff in 1910 (4), which list the associations spoken to each of 100 stimulus words by a population of 1,000 subjects, including 500 hospital workers, 200 professional persons and 150 high school students. Since the two sets of data differ with respect to population of subjects, time when measured, and mode of response (written or spoken), an estimate of the correlation between base probability and base associative probability based on these tables cannot be expected to approach unity even in the case of stochastic equivalence.

The Kent-Rosanoff tables list each combination of letters as a separate word, but the published version of the magazine count combines the entries for certain inflected forms (e.g., plurals ending in *s*, verb forms ending in *ed* or *ing*) with the root forms. The corresponding entries in the Kent-Rosanoff tables have been combined in the present analysis to make the two sets of frequencies comparable.

The stimulus words used in the Kent-Rosanoff study constitute a very unrepresentative sample of the set of all possible words. All are nouns and adjectives, for example, and all are words of medium frequency (their median log frequency is 2.6 with a standard deviation of 0.5 log unit). Weighting each associative frequency by the magazine-count frequency of the stimulus word appears to be of limited value in such a case. On the other hand, the weighting procedure greatly increases computational labor. Computation of base associative probabilities has therefore been simplified by using the approximation

$$\sum_i P(w_i)p_{si}(r_j) \simeq 0.01 \sum_i p_{si}(r_j) \quad [11]$$

in which the 100 stimulus words receive equal weights.

The notation may be simplified somewhat for the purpose of describing the data by dropping the indicators w , s , and r which distinguish the various functions of words in the experiments. The base probability is then denoted by $P(j)$ and the base associative probability by $\sum_i P(i)p_i(j)$. The empirical estimates

of these theoretical quantities will be referred to as the *magazine-count frequency* $F(j)$ and the *summed Kent-Rosanoff frequency* $\Sigma f_i(j)$, respectively.

ANALYSIS OF DATA

Figure 1 is a plot of the relation between magazine-count frequency and summed Kent-Rosanoff frequency on logarithmic coordinates. The numbers bordering the scatter plot represent the numbers of words having frequencies of zero, whose logarithms are therefore negative infinite. A significant positive correlation is evident from the figure, but the relationship is by no means rectilinear over the entire range. The form of the regression lines, as well as the magnitude of the correlation, is therefore of interest. Both types of analysis, in turn, require that the form of the conditional distribution functions be known.

Variation of base associative probability with base probability constant. Consider first the conditional distributions in the vertical dimension of Fig. 1. These distributions show what proportion of words having a given magazine-count frequency occur 0, 1, 2, ... times as responses in the Kent-Rosanoff tables. Figure 2-A presents several such distributions on logarithmic-normal coordinates. A straight line on these coordinates indicates that the logarithms of the variable plotted on the abscissa are normally distributed with a mean indicated by the intercept of the line and a standard deviation inversely proportional to its slope. The sets of points in Fig. 2-A are well described by a set of lines that are both straight and parallel, indicating that the conditional distributions of log base associative probability for given base probability are both normal and homoscedastic. (Actually, there is a slight increase in the standard deviation of $\log \Sigma f_i(j)$ at the highest magazine-count frequencies, but this is due to the use of wider class-intervals in that range in order to obtain distributions of more than 100 cases.)

The regression of summed Kent-Rosanoff frequency on magazine-count frequency is shown by the solid circles in Fig. 3-A. The coordinates are logarithmic. Medians have been substituted for means in the regression analysis because the distributions include zero frequencies the logarithms of which are infinite. The regression line has a maximum at

a magazine-count frequency of about 800 and falls off sharply on both sides of that point.

It may be observed that the regression line in Fig. 3-A covers only about half the total range of $\log F(j)$. Words that are rare in the magazine count tend not to occur at all in the Kent-Rosanoff tables, which comprise a much smaller sample. When the magazine-count frequency is small (< 150), the median summed Kent-Rosanoff frequency is therefore zero and the median $\log \Sigma f_i(j)$ is negative infinite. At the very highest magazine-count frequencies more than half the words again fail to occur in the Kent-Rosanoff tables, so that the median $\log \Sigma f_i(j)$ is negative infinite in that range too.

An estimate of the regression line over the full range can be found by using a theorem which states that the line connecting a given fractile of each of a set of normal, homoscedastic distributions is parallel to a line connecting any other fractile of the same set of distributions.² Since the highest fractiles of the conditional distributions of $\log \Sigma f_i(j)$ are finite for all magazine-count frequencies, the slope of the regression line can be estimated from the slope of a line connecting one of those fractiles. In Fig. 3-A the 95th and 75th percentiles of the $\log \Sigma f_i(j)$ distributions are indicated by crosses and triangles, respectively. These data indicate a regression line that is rectilinear and positive for magazine-count frequencies of less than 800 ($\log F(j) < 2.9$). (The leveling-off of the 95th percentiles for $\log F(j) < 1$ is due to an artifact explained in the following paragraph.) Above that point the regression line undergoes a sudden reversal until at the highest magazine-count frequencies its slope approaches³ $-\infty$. The significance of this reversal will be taken up in a later section.

The leveling-off of the 95th percentiles for $\log F(j) < 1$ is due to the omission of a number of words in that range. The magazine count is

² A proof may be indicated for a set of normal distribution functions $G_1(x), G_2(x), \dots, G_n(x)$ having standard deviations $\sigma_1 = \sigma_2 = \dots = \sigma_n = \sigma$. Let a and b be two proportions, $0 \leq a < b \leq 1$, for which the fractiles of the various distribution functions are $x_1(a), x_2(a), \dots, x_n(a)$ and $x_1(b), x_2(b), \dots, x_n(b)$, respectively. If $h(a)$ and $h(b)$ are the corresponding fractiles of the standardized normal distribution function, we can write $x_1(b) - x_1(a) = [h(b) - h(a)]\sigma = x_2(b) - x_2(a) = x_n(b) - x_n(a)$. A line connecting the points $x_1(a), x_2(a), \dots, x_n(a)$ will therefore lie $[h(b) - h(a)]\sigma$ units of x below a line connecting the points $x_1(b), x_2(b), \dots, x_n(b)$. Hence the two lines will be parallel.

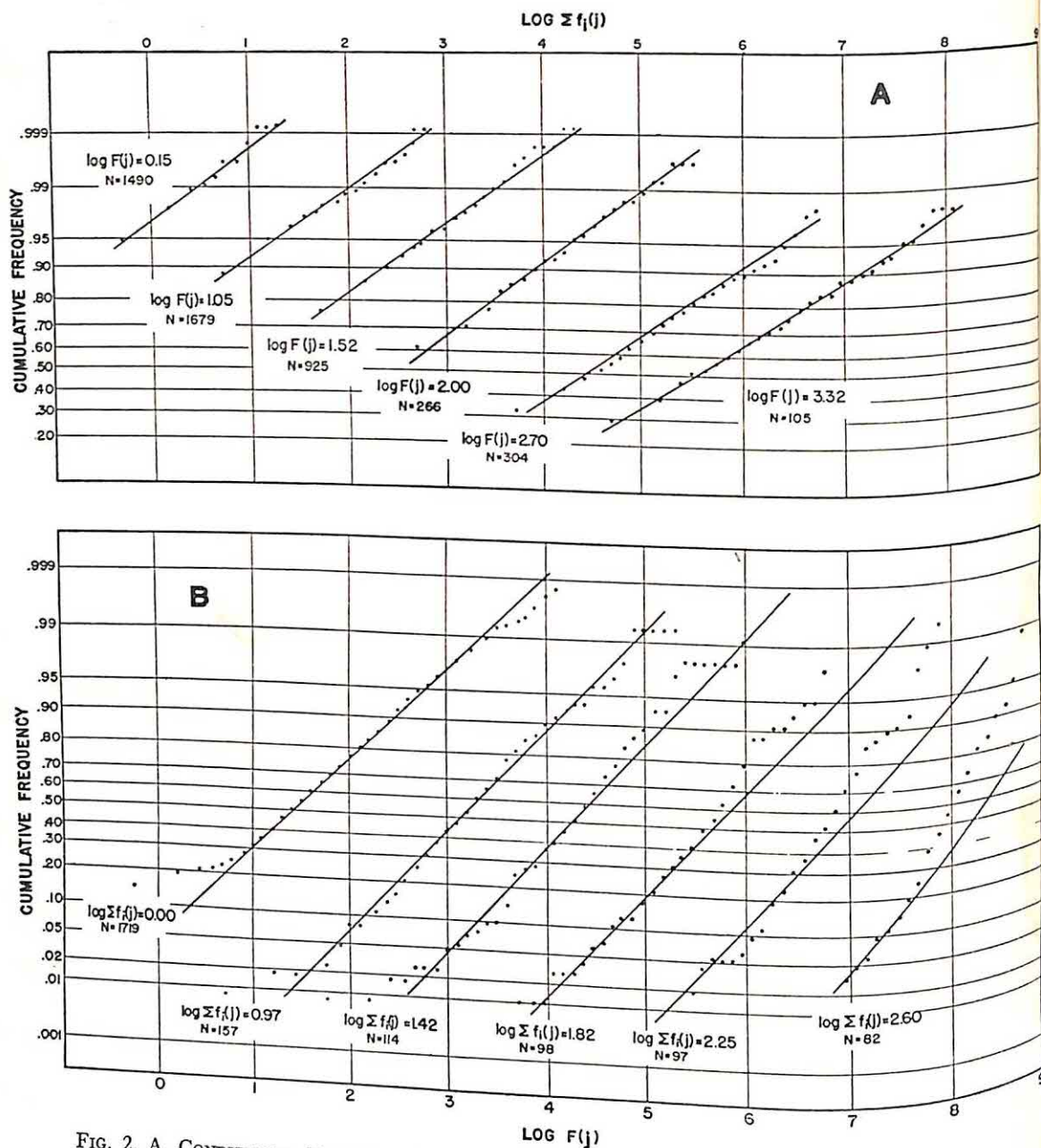


FIG. 2. A. CONDITIONAL DISTRIBUTIONS OF SUMMED KENT-ROSANOFF FREQUENCY FOR SELECTED VALUES OF MAGAZINE-COUNT FREQUENCY. B. CONDITIONAL DISTRIBUTIONS OF MAGAZINE-COUNT FREQUENCY FOR SELECTED VALUES OF SUMMED KENT-ROSANOFF FREQUENCY. Ordinate is proportional to the normal deviate, abscissa is logarithmic. The distributions have been separated by displacing each one to the right one log unit.

published as one of four word counts in the Thorndike-Lorge tables (8). These tables do not list the frequency of a word in each count if the combined frequency of the word in all four counts is less than 18. Moreover, the combined frequency itself is given only if it is greater than 3. In the data of Fig. 1 only magazine-count frequencies specifically listed in the tables are used. Words with combined

frequencies of less than 18 are thus assigned a magazine-count frequency of zero in Fig. 1 even though they may actually have occurred up to 17 times in the magazine count.

The number of these words actually having a given magazine-count frequency can be estimated approximately by assuming that the occurrences making up the combined frequency of a word are distributed randomly

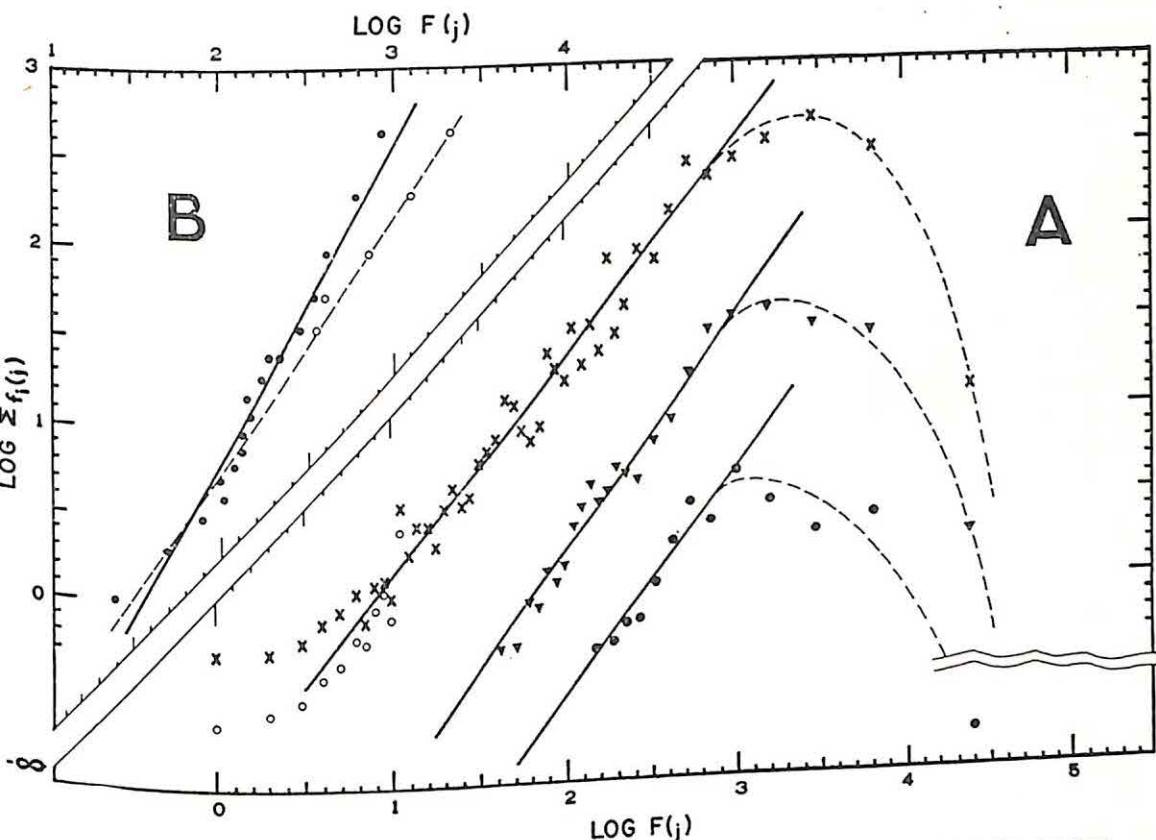


FIG. 3. A. REGRESSION OF SUMMED KENT-ROSANOFF FREQUENCY ON MAGAZINE-COUNT FREQUENCY.
 B. REGRESSION OF MAGAZINE-COUNT FREQUENCY ON SUMMED KENT-ROSANOFF FREQUENCY
 Logarithmic coordinates. Solid circles, triangles, and crosses represent 50th, 75th, and 95th percentiles, respectively. Open circles indicate corrected data as described in text.

among the four equal component counts.³ The summed Kent-Rosanoff frequency can be found for each word with a known combined frequency and assigned on a proportionate

³ Let Q be the combined frequency of a word in all four counts and q_1, q_2, q_3 , and q_4 the unknown component frequencies, any one of which may be chosen to represent the magazine-count frequency. Then $\Sigma q_i = Q$ by definition and the limits $4 \leq Q \leq 17$ and $0 \leq q_i \leq 17$ follow from the nature of the Thorndike-Lorge tables. The number of possible arrangements of the four component frequencies that satisfy these conditions for given Q may be denoted by A_Q , and the number of these arrangements that assume a frequency m in the component representing the magazine count, $A_{m|Q}$. The probability that a word with a combined frequency Q has a magazine-count frequency m may be estimated from the ratio $A_{m|Q}/A_Q$. Suppose N_Q words are listed with combined frequencies of Q in the Thorndike-Lorge tables. Then $n_{m|Q} = N_Q(A_{m|Q}/A_Q)$ estimates the number of words with combined frequencies of Q that have magazine-count frequencies of m . The estimated number of words with magazine-count frequency m that are omitted from the published Thorndike-Lorge tables is therefore $n_m = \Sigma_Q n_{m|Q} = \Sigma_Q N_Q(A_{m|Q}/A_Q)$.

basis to the estimated magazine-count frequencies. In this way the words omitted from the published tables can be assigned values of $F(j)$ and $\Sigma f_i(j)$. These data can be used to correct Fig. 1.

The 95th percentiles of the $\Sigma f_i(j)$ distributions for these corrected data are shown by the open circles in Fig. 3-A (no correction is indicated for magazine-count frequencies of 12-17 where the change was not appreciable). The corrected data fall quite close to the theoretical straight line except for magazine-count frequencies of three or less ($\log F(j) \leq 0.5$). In that range the correction is only partial because the published tables do not report words with combined frequencies of less than 4. The partial correction has a large enough effect, however, to indicate that complete correction would eliminate the remaining deviations of these points from the rectilinear regression line.

Variation of base probability with base associative probability constant. Consider next the con-

ditional distributions in the horizontal dimension of Fig. 1, which show the proportion of words with a given frequency in the Kent-Rosanoff tables that occur 0, 1, 2, ... times in the magazine count. Figure 2-B shows several of these distributions on logarithmic-normal coordinates. In this case the sets of points do not describe parallel straight lines as in Fig. 2-A. Two artifacts, however, affect these data: the omission of words with magazine-count frequencies of less than 18, described in the preceding paragraphs; and the reversal of the regression line for $\log \Sigma f_i(j)$ on $\log F(j)$, shown in Fig. 3-A. When the effects of these two factors are taken into account, the theoretical distributions of log base probability, indicated by the straight lines in Fig. 2-B, turn out to be normal and homoscedastic.

The first factor is responsible for the leveling-off at the lower ends of the distributions of Fig. 2-B (the effect does not show up in the two right-hand distributions because they contain no points in the affected range). It has been pointed out that words with combined frequencies of less than 18 in the four word counts of the Thorndike-Lorge tables had to be assigned a frequency of zero in Fig. 1 although their true magazine-count frequencies may be as high as 17. This produces an excess of words in the interval $F(j) \leq 0.5$ and a corresponding deficiency in the interval $0.5 < F(j) \leq 17.5$. In the cumulative plot of Fig. 2-B, therefore, the points for $F(j) = 0.5$ (-0.3 on the logarithmic abscissa) lie well above their true values, while those for $F(j) = 17.5$ ($+1.2$ on the logarithmic abscissa) are correct. The correction procedure described in connection with Fig. 3-A eliminates this type of deviation altogether.

The second factor, the reversal of the regression line of Fig. 3-A, is responsible for the deviations at the upper ends of the distributions of Fig. 2-B. The effect of the reversal is to lower the positions of points at the extreme right of Fig. 1 from the positions they would occupy if the regression line of Fig. 3-A were rectilinear over its entire range. At $\log F(j) = 3.0$ the median point is lowered about 0.1 log unit; at $\log F(j) = 3.5$, about 1 log unit; and at $\log F(j) = 4.0$, about 2 log units. Consequently, a horizontal slice through the upper part of Fig. 1 has a large proportion of its points in the range $\log F(j) \geq 3.0$ removed by the reversal of the regression line. A dis-

tribution of $\log F(j)$ at this level therefore has the form of a truncated normal distribution (2, §6.9). Distributions of $\log F(j)$ taken at successively lower levels of Fig. 1 have proportionately fewer points removed by the reversal. The degree of truncation should decrease accordingly. At the very lowest levels of Fig. 1, in fact, the reversal actually adds more points from higher levels than it removes to still lower levels. The net effect at these levels is thus to add a small number of points in the affected range.

In Fig. 2-B, distributions of $\log F(j)$ for successively lower levels of $\log \Sigma f_i(j)$ are arranged from right to left. All six sets of data can be brought into accord with the theoretical straight lines by replotting them for the following assumed degrees of truncation (reading from right to left): 55%, 35%, 18%, 9%, 0%, and -10% (negative truncation indicates the addition of points instead of their removal). For each set of data the degree of truncation that gives the best fit to the theoretical line is of the order of magnitude required by the data of Fig. 1.

The medians of both empirical and theoretical distributions are used in Fig. 3-B to indicate the regression of magazine-count frequency on summed Kent-Rosanoff frequency. Solid circles represent medians of the empirical (truncated) distributions, open circles the medians of the theoretical (normal and homoscedastic) distributions. (For $\log \Sigma f_i(j) \leq 1.3$ the solid circles represent both sets of medians.) Each set of data is well described by a straight line, the slope for the empirical medians (solid line in Fig. 3-B) being slightly steeper than the slope for the theoretical medians (dashed line).

Since the medians of the theoretical distributions have to be estimated from the theoretical lines of Fig. 2-B, their accuracy is open to question. As a test of their validity, the slope of the line fitted to the theoretical medians can be estimated by a second procedure. The truncation of the empirical distributions of $\log F(j)$ does not extend to the lowest fractiles. Hence a line passed through a low fractile of the empirical distributions should have the same slope as the dashed line of Fig. 3-B representing the medians of the theoretical distributions.² The line of best fit to the 25th percentiles of the empirical distributions has a slope of 1.45 ± 0.1 , compared to 1.40 ± 0.1 for the dashed line of Fig. 3-B.

By way of comparison, the dashed line of Fig. 3-B, representing the medians of the empirical distributions, has a slope of 1.85 ± 0.1 .

Correlation between base probability and base associative probability. Only a small proportion of the words represented in Fig. 1—perhaps 0.5%—are involved in the reversal of the regression line in Fig. 3-A. The behavior of all other words is described by the rectilinear regression lines of Fig. 3-A and 3-B (dashed line). For these data, which satisfy the assumptions of a normal correlation surface (2, p. 585), the product-moment correlation coefficient ρ can be used to measure the degree of association between base probability and base associative probability. Numerical computation of ρ is impossible because of the presence of infinite scores. The correlation can be determined from the angle θ between the two regression lines, however, since

$$\tan \theta = \frac{(1 - \rho^2)\sigma_1\sigma_2}{\rho(\sigma_1^2 + \sigma_2^2)}.$$

(2, Equation 19.7.12). For the present data, θ is 3.5 degrees (Figs. 3-A and 3-B), σ_1 is 1.10 log units (Fig. 2-A), and σ_2 is 0.75 log unit (Fig. 2-B). The corresponding estimate of ρ is $+0.937 \pm 0.001$.

This correlation must be interpreted in the light of several adverse factors affecting the data: the difference of 30 years between the dates of the two sets of data; the different composition of the two populations of subjects; the contrasts between the written language of the magazine count and the spoken language of the Kent-Rosanoff experiment; the small number of stimulus words tested in the Kent-Rosanoff study; and the use of the approximation to base associative probability instead of its exact definition. All these factors can be expected to introduce uncorrelated variance into the results. The true correlation between base probability and base associative probability, uncontaminated by these sources of error, must therefore be close to 1.0. Stochastic equivalence thus appears to apply to the words considered in this section.

The reversal of regression at high magazine-count frequencies. The words having the highest magazine-count frequencies are the ones that are responsible for the inversion in the regression line of Fig. 3-A. Table 1 lists the 25 words having the highest magazine-count

TABLE 1
SUMMED ASSOCIATIVE FREQUENCIES FOR THE 25 WORD CLASSES WITH THE HIGHEST MAGAZINE-COUNT FREQUENCIES

Word	$F(j)$	$\Sigma f_i(j)$	
	Magazine Count $N = 4,591,125$	Kent-Rosanoff $N = 100,000$	Jenkins $N = 19,100$
1. the	236472	0	28
2. and	138672	0	54
3. a	117222	0	12
4. to	115358	0	126
5. of	112601	0	38
6. I	89489	6	84
7. in	75253	0	29
8. was	58732	0	44
9. that	55667	0	73
10. it	52107	0	196
11. he	49268	0	82
12. you	42581	13	219
13. for	39363	6	54
14. had	34341	0	42
15. is	33404	0	135
16. with	32903	0	42
17. her	31824	1	60
18. she	31087	0	70
19. his	30748	0	104
20. as	30693	0	10
21. on	30224	1	64
22. at	26250	0	61
23. have	24456	19	47
24. but	23704	0	27
25. me	23364	14	242
Sum	1,535,783	54	1943
Sum/ N	0.33451	0.00054	0.10171

frequencies along with their summed Kent-Rosanoff frequencies. Without exception, the high magazine-count frequencies of these words are matched with very low summed Kent-Rosanoff frequencies. The average summed Kent-Rosanoff frequency is only 0.2% of the average magazine-count frequency, after correction for the difference in size of sample.

Study of Table 1 discloses that all the words listed there are of a characteristic type. They are words whose primary function is to connect other words in continuous discourse: articles, conjunctions, auxiliaries, prepositions, pronouns. These may be called *interstitial* words. The low frequencies of such words in the Kent-Rosanoff tables could be expected on two grounds. First, the subjects in the association experiment are instructed to emit only one-word utterances. The absence of continuous discourse eliminates the principal function of interstitial words, thereby reducing the number of occasions on which they are called forth. Second, it has been well established that sub-

jects tend to give associations of the same part of speech as the stimulus words (10, p. 347). The complete absence of interstitials among the 100 Kent-Rosanoff stimulus words would therefore result in low frequencies for those words as associations. The former interpretation of the result implies that interstitial words represent a special type of word to which the property of stochastic equivalence does not apply, while the latter interpretation holds that the exceptional behavior of these words is merely an artifact of the particular association experiment that has been analyzed.

The best evidence on the question is found in a recent association experiment by Jenkins⁴ in which over 50 interstitials, including all but three of the words in Table 1 (*in, her, as*), were among the 200 stimulus words. Table 1 shows the frequencies, summed over all 200 stimulus words, with which the words listed there occurred as responses in the Jenkins study. The effect is striking: there is a 200 to 1 increase in frequency over the Kent-Rosanoff experiment. The 25 words of Table 1, while they constitute only 0.05% of the Kent-Rosanoff responses, make up 10% of the Jenkins responses and 33% of the magazine count. Inclusion of interstitial words in the list of stimulus words thus goes a long way toward closing the gap between the summed associative frequencies and the magazine-count frequencies of interstitial words. Use of the complete formula for base associative probability instead of the approximation would narrow the gap even more.

The Jenkins experiment, based on fewer than 100 subjects, is not extensive enough to warrant a more elaborate analysis. But it shows clearly that the particular selection of stimulus words used by Kent and Rosanoff is chiefly responsible for the reversal of the regression line in Fig. 3-A. The sharpness of the contrast between the Jenkins results and those of Kent and Rosanoff further suggests the possibility that stochastic equivalence may apply to interstitial words as well as to noninterstitial ones. A large and carefully designed experi-

ment would be required, however, in order to test this possibility adequately.

SUMMARY AND CONCLUSIONS

The summed associative probability of a word, measured by its total frequency in the Kent-Rosanoff tables, is considered in relation to the probability of emission of the word in general linguistic usage, measured by its frequency in the Lorge magazine count. Two features stand out in the data: a high positive correlation (.94) for magazine-count frequencies of less than 800; and a sharp reversal in this relationship for higher magazine-count frequencies. Perhaps neither result, taken singly, is as striking as the fact that both appear so clearly in the same set of data.

The inverse relationship for high magazine-count frequencies is produced by a fairly small number of words of a special type, called *interstitial* words because their chief function is to connect other words in continuous discourse (i.e., conjunctions, articles, prepositions, auxiliaries, pronouns). Such words were systematically excluded from the list of stimulus words used by Kent and Rosanoff. As a consequence, they were rarely emitted as associations despite their high magazine-count frequencies. Data from another association experiment, in which interstitial stimulus words were included, indicate that the reversal of correlation found in the Kent-Rosanoff data can largely, if not completely, be eliminated by an unbiased selection of stimulus words.

The high correlation found for all other words indicates that the average probability that a given word will be emitted as a response in the word-association experiment is the same as its probability in general discourse. The language of association can therefore be said to be *stochastically equivalent* to the language of general discourse. This relationship implies that a person's selection of words is not affected by the artificial laboratory situation required for the association experiment. Hence the effect of a given stimulus word upon a subject's response must be due to a redistribution of the set of factors governing emission of words in general discourse. The results thus serve to validate the association experiment as a method for studying under laboratory control the processes governing the emission of words in general discourse.

⁴ I wish to thank Dr. J. J. Jenkins of the University of Minnesota for generously placing his data at my disposal and Miss Evelyn Segal for preparing the computations.

REFERENCES

1. COFER, C. N., & SHEVITZ, R. Word-association as a function of word-frequency. *Amer. J. Psychol.*, 1952, **65**, 75-79.
2. HALD, A. *Statistical theory with engineering applications*. New York: Wiley, 1952.
3. HOWES, D., & OSGOOD, C. E. On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.*, 1954, **67**, 241-258.
4. KENT, G. H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, **67**, 37-96; 317-390.
5. O'NEIL, W. M. The effect of verbal association on tachistoscopic recognition. Univer. of Minnesota Stud. on Verbal Behavior, Tech. Rep. No. 4, 1953.
6. RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287-293.
7. SKINNER, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, **1**, 71-76.
8. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia Univer., 1944.
9. THUMB, A., & MARBE, K. *Experimentelle Untersuchungen über die psychologischen Grundlagen der sprachlichen Analogiebildung*. Leipzig: Engelmann, 1901.
10. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
11. ZIFF, G. K. *Human behavior and the principle of least effort*. Cambridge, Mass.: Addison-Wesley, 1949.

Received March 12, 1956.

LOOKING AT OCCUPATIONS¹

WILLA FREEMAN GRUNES

University of California, Berkeley²

How does the young American look at jobs? What groups of jobs appear in his psychological field? What do these groups look like to the perceiving individual? Do different social classes see these things differently? These are phenomenological questions, of the sort MacLeod (9) meant when he suggested that the phenomenological method, found so fruitful in approaching classical problems in perception, be applied also in the field of social psychology. They concern the social world which is psychologically "there" for the subject, quite apart from what the experts might say is, or ought to be, there.

There have been few studies that attempt to answer such questions about aspects of the social world by attempting to discover perceived units and their characteristics. Methods for such investigation are woefully lacking and difficult to invent. The present study is presented as suggestive of new ways for mapping perceived social worlds.

If we can determine what people see, we can better understand what they think and do. Specifically, in the realm of occupations, we can better see why they make their particular vocational choices, and why too many choose certain types of overcrowded fields while too few choose others.

Occupational perceptions were selected for this study partly because they seem important for revealing the psychological meaning of social class memberships. Such meaning has long been neglected in research on class in America. These perceptions also offer a meeting place between psychology and economics and hence offer a promising area for mutual enrichment of both fields.

METHOD

Sample

About 150 students in each of eight high schools were tested in 1952-53. Four Northern California schools were included: Oakland, McClymonds, San Lorenzo, and Hayward. There were also two Southeastern schools: Talledega in Alabama and Orangeburg in South Carolina; one Northwest Central school: Bismarck in North Dakota; and one Northeastern: Clifton Heights in Pennsylvania. The communities of the school vary widely in size, social class, race, and in degree and type of industrialization. The diversity seemed desirable in order to assess the generality of the findings.

Procedure

Two group-administered devices, the Grouping Test and the Pick-A-Job Test, were developed and pretested for purposes of revealing the way perceived occupations tend to be grouped by high school students; i.e., the perceived structure of the world of jobs. Both were indirectly inspired by various studies by Krech [(8) and recent unpublished work] on the way rats and also men structure the world around them. In order to make it easy and threatless for subjects to reveal their perceptions, both devices are disguised on purpose. The subject is provided with an incomplete situation, out of which he must make some sense of his own in order to deal with it. Such a disguised and unstructured approach seems appropriate to an area of content in which subjects may be unable or unwilling to verbalize what they perceive. Through these two devices, a perceptual structure was sought which could be regarded as common to most high school students. The assumption was nevertheless avoided that all social classes—or all individuals—possess a similar structure, or that any perceived structure necessarily reflects objective social reality. In this respect the method differs from that used by Lloyd Warner and his associates (10) to study class structure.

Grouping Test. On the Grouping Test, the subject was presented with a list of 51 varied occupations, selected to represent each of the major categories in the Dictionary of Occupational Titles (11). He was given the following written instructions (somewhat abridged here) along with some oral clarification:

Let us see how fast you can think about various occupations. Group the following occupations as many ways as you can. For example, make groups according to what kinds of people work at the jobs, what kinds of work they do, etc. Give each group a title that tells what kinds of jobs belong under it. . . . At least two jobs must be in each group you make. You may use the same jobs under as many titles as you wish. For example: People who would probably know how to give

¹ The material discussed is adapted from a doctoral dissertation (6) done under the guidance of Professor Mason Haire. Especially extensive assistance was given by H. D. Mugass, R. R. Lange, W. C. Rasmussen, Dorothy W. Gottlieb, Katherine M. Cavanagh, Leon Festinger, and Camille J. Miller, as well as by the administrators and their staffs of the schools involved.

² The author now resides in Mandan, North Dakota

first aid: airline stewardesses, doctors, nurses, physical education teachers. . . .

The more groups you make within the time limit (15 minutes) the higher your score. You can use any types of groups you can think of. . . . Be sure to include all of the occupations on the list that belong under each of your titles, because you also get a higher score for every job you list that belongs there. Here is the list of occupations. . . .

The resulting data were tabulated and coded to indicate the jobs which tend to be placed together in the same groups, the grouping categories reflected in the titles, and the jobs most frequently placed under each type of title. A set of job clusters common to most students' view of the world of occupations was discovered. Content analysis of the titles was done by two coders working independently, using only categories which yielded at least 80 per cent agreement.

Since completing the present study, the author has learned of a similar effort by Campbell (1) to develop a method for establishing subjective occupational groupings. His study failed, however, to reveal any common job clusters. The present method differs from his in permitting the subject to use each occupation in as many groups as he wished and to make as many (overlapping) groups as he could, as well as in the procedures used to analyze the data. These provisions appear more appropriate than Campbell's to the presumably complex nature of the perceived structure of the occupational world.

Pick-A-Job Test. The Pick-A-Job Test offered the subject brief descriptions of six different people and asked him to assume the role of vocational counsellor and to suggest five appropriate jobs for each one. For example, the first description appeared thus:

- A. Man
 1. _____
 good-looking 2. _____
 clean 3. _____
 well-educated 4. _____
 self-confident 5. _____

Results were analyzed to indicate the typical sorts of jobs suggested for each description.

The personality descriptions used on the Pick-A-Job Test were derived from earlier use of what may be called the Haire Describe-A-Person method, a modification of a procedure originally developed by Haire and described by Haire and Grunes (7). High school students were given a few rather neutral facts about a fictitious person, and also his occupation, and were asked to guess the sort of person he was. Their guesses were coded for content and compared with descriptions obtained from similar subjects when the occupation was omitted from the given facts. Differences between the two sets of descriptions revealed stereotypes associated with the occupations selected for study. The Pick-A-Job Test was based on the typical descriptions thus obtained of businessman (Description 1 above), ditch-digger, actress, and farmer. Additional descriptions based on the typical businessman and ditch-digger were included, with the characteristics ascribed in each case to a woman.

Besides these procedures, the Gough Home Index (4), a measure of relative social status of high school students, was administered. It yields a continuous dis-

TABLE 1
 SOCIAL CLASS DIFFERENCES IN FREQUENCY OF
 GROUPING OTHER OCCUPATIONS WITH DOCTOR
 (Entries represent percentage of groups involving
 "doctor" made by each social class)

Co-Member	Home-Index Social Class		
	Lower	Middle	Upper
Nurse	31	66	53
College Professor	38	59	80
Teacher	44	56	73
Social Worker	13	24	47
Air Line Stewardess	56	38	33

tribution of scores, but was here employed to classify subjects into three status groups (low, middle, and high) by cutting the distribution for each school at the two points where the large central frequencies began to fall off.

RESULTS

Deriving the Job Clusters

Data were compiled on each job listed on the Grouping Test, showing, for each social class, the other jobs with which a given job was grouped and the types of titles under which it was listed.³ As a sample, the following information was gathered about the "doctor":

In at least 50 per cent of the instances in which doctor was listed, the following jobs appeared in the same group: nurse, college professor, and teacher. In at least 40 per cent, the engineer and minister were in the same group.

The dish-washer, ditch-digger, farmer's wife, hired man on farm, rancher, and janitor were in the same group with doctor less than 5 per cent of the times it appeared.

Doctor appeared in at least 50 per cent of the groups in which each of the following jobs appeared: nurse, college professor, teacher, engineer, social worker, reporter, businessman, and business-woman.

Table 1 shows the percentage of groups made by each social class that included both the doctor and each of the five occupations listed. Note that the higher status subjects tended to group doctor with other professionals more consistently than did the lower status subjects.

The categories of Grouping Test titles under which doctor was listed at least 50 per cent of the times they occurred include: educational barriers, much education, good looking, lot of money, personality barriers, social skills, ability barriers, brains, verbal skills, "skilled," "trained," training barriers.

Doctor was listed under less than 50 per cent of the titles categorized as follows: charm, strong, physical barriers, little or no education, "common labor," dumb, not own boss, little money.

³ All the data are available in tabular form in the author's dissertation (6).

TABLE 2
A SAMPLE JOB CLUSTER MATRIX: CLUSTER A

	Job Title									
	Doc.	Nur.	Coll.	Teach.	Eng.	Min.	Social	News.	Busm.	Busw.
Co-Members										
Doctor	X	++	++	++	++	++	++	++	++	++
Nurse	++	X		++	+	++	++	+		+
College Professor	++	++	X	++	++	++	+	++	++	++
Teacher	++	++	++	X	++	++	++	++	+	++
Engineer	+				X	+				
Minister	+	++	++	+	+	X	++	+		+
Social Worker					+	+	X			
News Reporter						+	+	X		+
Businessman			+	+		+	+	+	X	++
Business Woman			+			+	+	+	++	X
Air Line Stewardess		+				+	+	+		+
Phys. Ed. Teacher					+	+	++	etc.*	etc.*	etc.*
Nonmembers										
Dish Washer	-		-	-	-	-		-	-	-
Ditch Digger	-	-	-	-	-	-		-	-	-
Farmer's Wife	-		-	-	-	-		-	-	-
Hired Man, Farm	-		-	-	-	-		-	-	-
Rancher	-		-	-	-	-		-	-	-
Married Woman			-	-	-	-		-	-	-
Welder			-	-	-	-		-	-	-
Truck Driver			-	-	-	-		-	-	-
Maid			-	-	-	-		-	-	-
Housewife			-	-	-	-		-	-	-
Janitor	-		-	-	-	-		-	-	-
Garbage Collector			-	-	-	-		-	-	-
Factory Worker			-	-	-	-		-	-	-
Farmer			-	-	-	-		-	-	-
Cleaning Woman			-	-	-	-		-	-	-

Note.—++ means that 50 per cent or more of the times that the occupation at the top of the column was listed, it appeared in the same group as the occupation at the left of the row.

+ means that such co-grouping occurred in at least 40 per cent, but not as many as 50 per cent of the total listings of the occupation at the top.

— means that the occupation at the left appeared in fewer than 5 per cent of the groups in which the occupation at the top occurred.

* In the case of news reporter, businessman, and business-woman, associations were found with additional co-members not listed here. Data for the other occupations are complete.

From the above data, it was predicted that the first Pick-A-Job description, as given above, would elicit the suggestion of doctor, among others. Actually, although the students had the whole world of occupations from which to choose, and five different choices to make, doctor accounted for fully 10 per cent of all jobs chosen for that description.

Similar data for each job were combined to define and characterize the commonly perceived job clusters. Table 2 gives, for example, the matrix of relationships which led to the definition of the cluster (Cluster A) to which the doctor belongs. The data presented also provide some evidence that Cluster A overlaps with another cluster.

A complete matrix, similar to Table 2 but including all 51 jobs on the Grouping Test, was used in determining the commonly perceived clusters.⁴ Unless an occupation was grouped

⁴ The matrix composes Part IV of the Appendix of the dissertation, pp. 307-316 (6).

with at least one member of a cluster at least 40 per cent of the time, it was not considered to be even a borderline member of that cluster. To be considered a clear member, it had to be grouped at least 50 per cent of the time with most of the other clear members of the cluster.

A number of jobs were found to be borderline or clear members of two clusters, and these jobs were taken as evidence that the clusters overlap. No jobs were found to be even borderline members of more than two clusters.

Confirmation of the clusters thus derived was found in the "nonmembers," which always came from "distant" (nonoverlapping) clusters, and in the data from the Pick-A-Job Test. Students tended to choose jobs from a highly restricted set of clusters for each description on the Pick-A-Job Test. These clusters were found to be predictable from a knowledge of the cluster membership of the job the stereotype of which had been drawn upon in formulating the description. For example, Description 1 (given

above) elicited almost exclusively members of the cluster (A-B) to which "businessman" belongs and members of the two clusters (A and B) with which A-B overlaps.

Implicit criteria in grouping. Characteristics of the clusters were determined by an analysis of the type of titles under which the cluster members were most frequently grouped—or not grouped—on the Grouping Test. As shown by the types of titles they choose for their groups, students typically show consciousness of certain dimensions in thinking about jobs, most notably: barriers to attaining the jobs (especially educational barriers), status, skill and knowledge, and the nature of the work.

The empirical job clusters. The overlapping job clusters and the outstanding characteristics ascribed to them, as revealed by the Grouping Test and confirmed by the Pick-A-Job procedure, are as follows:

Cluster A. These jobs are characterized as requiring much education, ability, brains, social and verbal skill, and as yielding much money; they are almost never seen as requiring physical strength. Included from the list provided are college professor, minister, doctor, nurse, teacher, and less clearly engineer, social worker, reporter, and businessman.

Cluster A-B. These jobs have characteristics similar to those of Cluster A, but with much less stress on educational barriers. They include businessman (also a member of A), and business-woman (also a member of B).

Cluster B. These jobs are seen as "business" and as requiring verbal skills. They include business-woman, office-worker, secretary, stenographer, bank clerk, and, less clearly, salesman, saleswoman, and store clerk.

Cluster C. These jobs are seen as divided into two subclusters: C-1, characterized by skill and strength, and C-2, seen as hard work and "common labor" done by a worker who is strong and not his own boss. C-1 includes carpenter, mechanic, and welder; C-2 the farmer, rancher, hired man on a farm, truck-driver, and less clearly, ditch-digger.

Cluster C-D. These jobs are seen as involving hard work, "common labor," and as requiring little education or anything else except strength. The workers are seen as stupid and as not their own bosses. They include ditch-digger (also a member of C), garbage-collector, and janitor.

Cluster B-D. These jobs are seen as "business" and as requiring social skill. They include salespeople and store clerks (also in Cluster B) and also waiters and waitresses.

Cluster D. These jobs are seen as requiring little (not even strength), as "common labor," and hard work, and their holders as stupid people of little education who are not their own bosses. They include dishwasher, cook, maid, housewife, married woman who stays home and keeps house, cleaning woman and, less clearly, elevator operator, waitress (also a member of B-D), and janitor (also a member of C-D).

Certain occupations on the list could not be placed in the above system of interrelated clusters, including one rather weak group including actresses, airline stewardesses, and models, and a strong one including only policemen and firemen. Jobs which seemed to stand completely alone were factory worker (distantly related to C-D and D), aviator, musician, physical education teacher, farmer's wife, tailor, barber, foreman, railroad conductor, and baseball player. A still more exhaustive list of occupations might well reveal further relationships.

Prestige Hierarchy

In the characteristics ascribed to the clusters, the prestige hierarchy inherent in the common cognitive structure is revealed, with positive and barrier-type characteristics both tending to be concentrated in the A and A-B Clusters, and least in Cluster D and the clusters overlapping it. Physical strength is an interesting exception to this rule. This prestige hierarchy agrees well with the conclusions of other investigators (2, 3) who have asked various samples of Americans to rank lists of jobs according to their "social standing."

There is a large gap between the perception of professionals, business, and office people (Clusters A, A-B, and B, but not B-D) on the one hand, and manual and mechanical workers on the other, with only relatively rare cases occurring in which they are seen as members of the same group. Some of the evidence for this deep gap may be seen by studying the "non-members" in Table 2. All of these cases of extremely rare grouping with Cluster A occupations are from Clusters C, D, and C-D. The gaps between skilled and unskilled workers and between clerical and professional workers are not nearly so wide.

Stereotyped Job Choices

Students agree remarkably well on the occupations and clusters of occupations seen as appropriate for the people described on the Pick-A-Job Test. For each description, their choices are sharply limited to a few salient possibilities out of the thousands of occupations from which they were free to choose. In answer to Description 1, for example, the following job categories accounted for about 70 per cent of all the suggestions: salesman, manager, businessman, teacher, doctor, lawyer, and "office worker."

TABLE 3
OCCUPATIONS MOST COMMONLY GROUPED WITH
COLLEGE PROFESSOR ON THE GROUPING TEST BY
SOCIAL CLASS

Co-Member	Cluster	Percentage of groups which contain both the occupation listed and college professor		
		Lower Class	Middle Class	Upper Class
Doctor	A	53%	70%	75%
Minister	A	33	53	63
Teacher	A	73	68	81
Engineer	A	30	44	58
Businessman	A-B	50	47	44
Business-woman	A-B	60	43	42
Office worker	B	33	26	17

Note.—Few of the differences in percentage shown are statistically significant, but they seem important because of their consistency. Those for college professor are typical of those for all other Cluster A occupations. There were no relevant cases showing an opposite trend.

For Description 2—a man who is hard-working, strong, not very bright, and likes to follow orders—students mainly choose jobs from Clusters C and C-D, which contain mostly masculine occupations. The woman with the same traits is steered mainly toward Clusters B-D and D. For the other descriptions, however, the sex of the person described changes the particular jobs selected, but not the clusters from which they come.

Group Differences in Perception of Occupations

Class differences. Although there is considerable agreement among students about the common pattern of clusters, there are some social class differences in cognition of the occupational world. Most notably, students from the upper classes tend to make a sharper distinction between business and professional people than do the lower classes. The latter tend to lump Clusters A and A-B and B into one high-prestige cluster. They also tend to exclude from that cluster certain occupations often placed there by upper classes—such as the social worker, minister, engineer, reporter, and musician. Table 3 presents a sample of the evidence for these conclusions, using the occupations grouped with college professor on the Grouping Test as examples. Each value in the table represents the percentage of the total number of groups, made by the subjects in a particular social class, which contain both the occupation at the left and “college professor.” Note that the college professor is grouped with

other professionals more consistently, and with business people less frequently, as status rises. (See also the data on doctor listed above.) In general, the upper classes seem to be more finely differentiated about the upper occupational strata. There is also some less clear evidence that the lower classes are more differentiated about the lower prestige jobs.

The class differences in characteristics attributed to particular jobs follow lines predictable from the class differences in the job clusters themselves. Subjects from the lower classes tend more often to include jobs from Clusters A-B and B—as well as A, which all classes stress—when listing jobs that share certain positive and barrier-type characteristics, especially those concerning education and ability.

Certain particular occupations on the Grouping Test list were placed in different clusters by different classes. The clearest example is the “engineer,” interpreted as a professional by upper classes and as a skilled mechanical worker by lower.

These class differences are evident in coherent patterns in the data, but could not feasibly be subjected to statistical test. The group differences described below, except where otherwise noted, were found to be significant at the .05 level of confidence.

Subjects higher in social status appear to be relatively lacking in both status and barrier consciousness, but relatively strongly conscious of skill and knowledge, pace of work, and money. It may be that students are less aware of the job aspects about which they feel less insecure. The fact that students whose breadwinners are in manual or mechanical work are relatively less conscious of strength seems to support this inference.

The Pick-A-Job Test results reflect the greater tendency of the lower classes not to differentiate the top three job clusters. While higher status subjects restrict themselves closely to professionals and business people for both the male and female version of Description 1, clerical occupations are often selected by lower classes. For example, clerical work is suggested for the male version by 31.3 per cent of the lower-class subjects, by 15 per cent of the middle class, and by only 2.6 per cent of the upper class (difference significant at the .01 level). For Description 2, upper classes tend to

name more stereotypically low-prestige jobs like the garbage-collector and ditch-digger—perhaps again indicating their coarser distinctions among lower-strata occupations.

Sex differences. The study revealed some outstanding differences in the way boys and girls see the world of jobs. Through the titles they used on the Grouping Test, girls showed more awareness—or at least more articulateness—concerning all the common types of job aspects, with the exceptions of security, money, and hours of work. If the female is more status conscious, as has been claimed, she is also more conscious of a large number of other perceptual dimensions of jobs.

Sex differences were also found in the answers to Description 1 and its feminine counterpart on the Pick-A-Job Test, with the girls showing a stronger tendency than boys to name teaching jobs for both versions. Boys tend to select clerical rather than professional jobs for the female version, while girls choose mostly professional jobs, as do both sexes for the male version—perhaps a sex difference in willingness to limit capable females to lower social position.

Regional differences. In broad outlines, students in all eight schools see the same world of occupations. But consistent trends in the data indicate that those in the large industrialized communities tend, like higher status subjects, to see a relatively sharp professional business distinction. Students in the Southeastern schools appear, through their Grouping Test titles, to be more conscious of status and education than the others tested; but otherwise there is considerable regional similarity in the types and relative potencies of dimensions used for categorizing occupations. The social worker is the outstanding example of a job that is grouped differently in different regions: it is much more clearly professional in the large Northern cities.

Despite the class, sex, and regional differences mentioned, in general the students seem to agree rather closely in their basic perceptual structure concerning occupations.

DISCUSSION

The Grouping Test technique, in which a collection of items is set before the subject for him to record the groups or units that he perceives spontaneously, is easily adaptable to many problems of social perception. Instead of

occupations, one could list, for instance, names of political figures, nationalities, religions, brands of food, business corporations, or periodicals. The general method of the Pick-A-Job Test is likewise adaptable. Since it involves giving the subject hypothetical group characteristics and asking him to name members of the group, it could be called the Pick-The-Members technique. In other adaptations, subjects could be asked to name, instead of jobs, clubs to which the described individual might belong, public policies he might be likely to favor, movies he might enjoy, etc.

The Haire method (Describe-The-Person or Describe-The-Member), used here in conjunction with the Pick-A-Job Test, seems generally adaptable for yielding hypotheses about the characteristics of perceived groups. Together the two methods can be used to confirm Grouping Test data about the composition and characteristics of the perceived units. If the conclusions from the Grouping Test are valid, use of the Pick-The-Members technique on the same subjects should yield the members of the appropriate Grouping Test cluster, when that cluster's characteristics are used to form the description on the Pick-The-Members Test. And the Describe-The-Person method should tend to yield the characteristics found through the Grouping Test for the group of which the Person is seen as a member. The various methods were used in this manner in the present study to complement and validate each other.

Some of the implications of the study for vocational guidance are discussed elsewhere (5).

SUMMARY

A phenomenological study of the way American high school students perceive occupations is presented. A perceptual structure consisting mainly of seven overlapping job clusters is found to be characteristic of most subjects. Class and regional differences in the structure are analyzed and the different attributes ascribed to each cluster are examined. Two original types of methods are used which are adaptable to many other problems in social perception.

REFERENCES

1. CAMPBELL, J. D. Subjective aspects of occupational status. Unpublished doctor's dissertation, Harvard Univer., 1952.

2. COUNTS, G. S. Social status of occupations. *School. Rev.*, 1925, **33**, 16-27.
3. DEEG, M. E., & PATERSON, D. G. Changes in social status of occupations. *Occupations*, 1947, **25**, 205-208.
4. GOUGH, H. A short social status inventory. *J. educ. Psychol.*, 1949, **40**, 52-56.
5. GRUNES, WILLA FREEMAN. On perception of occupations. *Personnel Guid. J.*, 1956, **34**, 276-279.
6. GRUNES, WILLA FREEMAN. The American high school student looks at the world of occupations. Unpublished doctor's dissertation, Univer. of California, Berkeley, 1954.
7. HAIRE, M., & GRUNES, WILLA FREEMAN. Perceptual defenses: processes protecting an organized perception of another personality. *Hum. Relat.*, 1950, **3**, 403-412.
8. KRECHEVSKY, I. Hypotheses in rats. *Psychol. Rev.*, 1932, **6**, 516-532.
9. MACLEOD, R. B. The phenomenological approach to social psychology. *Psychol. Rev.*, 1947, **54**, 193-210.
10. WARNER, L., MEEKER, MARCHIA, & EELLS, K. *Social class in America*. Chicago: Science Research Associates, 1949.
11. Dictionary of occupational titles. (2nd ed.) Washington, D. C.: U. S. Bureau of Employment Security, 1949.

Received March 13, 1956

AGE DIFFERENCES IN THE PERCEPTION OF CLOSURE¹

HAROLD BASOWITZ AND SHELDON J. KORCHIN

Institute for Psychosomatic and Psychiatric Research and Training, Michael Reese Hospital

THE changes associated with aging are many and varied. Certain alterations in sensory acuity, speed of response, and performance on motor, memory, and selected intellectual tasks have been reasonably well established both by common observation and controlled experimentation (6, 8, 17). Yet the number of investigations concerned with more complex perceptual functioning in older persons has been surprisingly small. Of particular interest are comparative studies of Rorschach protocols typifying groups of differing ages and their implications for distinctive aspects of personality organization (3, 4, 5, 11, 18).

In the course of a larger research project dealing with psychological and neurophysiological correlates in aging, some provocative findings emerged related to a more circumscribed type of perceptual ability—that of closure. It is the purpose of this paper to consider this material. More specifically, the data to be presented involve the performance of two groups widely separated in age on two kinds of perceptual tasks which earlier factorial studies have claimed to reflect “the ability to fuse a perceptual field into a single percept” (21) and “flexibility of closure” (20). Within a broader context, this study is conceived as an attempt to enhance our general understanding of cognitive functioning in senescence. Related investigations are reported separately (14, 15).

METHOD

Two populations were sampled: a “young” group of 16 doctors and nurses composed of 5 men and 11 women, and an “old” group of 16 residents, 8 men and 8 women, in a home for the aged. The ages of the young ranged from 22 to 33, with a mean of 26.8 years; those of the old from 68 to 88, with a mean of 78.1 years. All Ss

were physically and mentally healthy and had adequate vision, using corrective lenses where necessary. Since one of the techniques to be discussed, Concealed Figures, has been found by other workers to be saturated with an intellectual factor (20), the groups were chosen so as to be roughly equated for intelligence. They were drawn from 24 young and 36 old persons used in the total study, to whom a battery of tests had been administered which included the Vocabulary subtest of the Wechsler-Bellevue Intelligence Scale (22) as well as the two perceptual tasks of concern in this paper. Matching for general intelligence was performed independently on the basis of vocabulary level, since it has been adequately demonstrated that this index declines relatively little in the normal process of aging (1, 2, 10, 22). The mean and *SD* of unweighted vocabulary scores for the young Ss so selected were 26.8 and 4.6; for the old, 28.6 and 4.8, respectively. Although these restricted groups enable us to focus most directly on influences of age uncontaminated by differences in intellectual ability, parallel analyses of the data have also been carried out for the larger parent populations. In all instances the age differences in performance are similar in direction and extent to those of the present study, at times being perhaps more extreme.

Two procedures were used:

Gestalt Completion Test. In this task, adapted by Thurstone (20) from a technique developed by Street (19), S is presented with a series of 72 ambiguous drawings which he is required to identify. Portions of the drawings are obliterated and at first view the material appears to consist of diverse and unrelated black blotches. Actually the fragments are arranged so as to suggest the outlines of distinct persons, objects, and/or situations. The accomplishment of closure involves organizing the seemingly scattered elements so that a compelling and unified whole emerges. In his factor-analytic studies of mechanical aptitude Thurstone (21) concluded that the test represented “the ability to fuse a perceptual field into a single percept.”

Following some examples, the instructions were: “When the signal is given, turn the page and identify some more pictures of the same kind. Work as rapidly as you can. *If some pictures are too difficult, skip them, and return to them later if you have time.*” Scoring criteria corresponded to those followed by Campbell and Zaks² in recent unpublished studies. Responses were scored 1 when the activities depicted or all of the persons and objects in the picture were accurately described. Half credit was allowed when partial integrations occurred and these were essentially correct.

Concealed Figures. Thurstone's (20) version of the Gottschaldt Figures Test (9) was employed. Forty-nine rows of figures are given, with a simple geometric design appearing at the left of each row followed by four more complex designs. The S identifies each of the more complex designs in which the simple figure is em-

¹ This study was supported by the U. S. Army under Contract No. DA-49-007-MD-261 through the Medical Research and Development Board, Office of the Surgeon General, Department of the Army, Washington, D. C. We gratefully acknowledge the aid of Mr. Ben Grossman, Director, and the staff of the Drexel Home, Chicago, in providing subjects and facilities. Dr. Albert Caron and Mrs. Ann Gunther were responsible for much of the data collection and Dr. Helen Heath for the statistical analyses.

² Personal communication.

TABLE 1
MEAN PERFORMANCES OF YOUNG AND OLD
GROUPS ON GESTALT COMPLETION
TEST

Group	N	Attempted	Right	Wrong	Omitted
Young	16	63.1	44.7	6.0	12.4
Old	16	60.7	15.5	15.9	29.4

bedded. Since the complex figure has "good gestalt" its properties of closure must be resisted as *S* searches for the smaller figure potentially hidden therein. The principal component of this test was termed by Thurstone (20) "flexibility of closure."

In essence the instructions stated: "In every row of designs the *figure* is concealed in at least one drawing and may be in as many as all four. Put a check mark (✓) in the parentheses below a drawing if it contains the *figure*. Put a zero (0) in the parentheses below a drawing if it does not contain the *figure*."

Both procedures were administered in booklet form with separate answer sheets. Testing occurred in individual sessions having a time limit of 10 minutes for each test. The instructions and illustrative examples on the front page of the booklets were read aloud and discussed to ensure that *S* fully understood the nature of the tasks.

RESULTS

During the 10-minute test period provided for Gestalt Completion, *S* undertook as many problems as possible, some responses being correct, some incorrect, and others omitted. In Table 1 are presented the mean frequencies of each type of response for each group. First to be noted is the lack of any appreciable difference between the number of items attempted by the young and old groups. "Attempted" is defined by the last item completed. Thus it appears that both groups at least examined equivalent portions of the total test in the time allotted.

Striking differences are found for the number of drawings correctly perceived. The young have a mean number right of 44.7, in contrast to 15.5 for the old. The *t* between these values is 7.68 which with 15 *df*—following Edwards' suggestion of halving the number of degrees of freedom in cases of unequal variance (7, p. 170)—is significant at $p < .001$. Equally sharp differences are seen on number of items incorrectly identified by each group, the mean for the young being 6.0, while that for the old is 15.9. These differences also are highly reliable ($t = 6.11$, $df = 30$, $p < .001$). Moreover, the old group has a significantly greater frequency of failures to respond, 29.4 omissions as com-

pared to 12.4 for young *Ss* ($t = 5.44$, $df = 30$, $p < .001$). "Omissions" are simply the number of items left blank of the total number attempted. It is therefore quite clear from these data that the ability of older persons to achieve perceptual closure is distinctly inferior to that of young individuals.

A qualitative analysis of wrong responses offers further insight into the characteristic functioning of the two groups. Almost invariably the types of errors committed by young *Ss* consist of inaccurate integrations of the entire picture. That is, the disjointed parts are fused into a whole, but the closure effected is not the intended figure. Sometimes the responses given depart but little from those credited as correct, and quite often the scorer can, as it were, perceive the figure as did *S*. By contrast, older persons usually, though less often, incorporate all the fragments, but in generally loosely organized and sometimes clearly distorted impressions. In addition, although occurring infrequently, all of the old exhibited one or more of the following features in their records: fixation on details of the drawings, perseveration of the same response on successive drawings, and a kind of passive acceptance of the elements as separates, with no genuine closure. In the last category are such answers as "rocks," "clouds," and "flock of birds," which imply a basic deficiency in organizing activity.

Table 2 describes the results for Concealed Figures. It is immediately apparent that the young group completed considerably more items on this test than the old. The difference far exceeds statistical significance ($t = 5.14$, $df = 30$, $p < .001$). However, even within the number of items completed the groups differ significantly in relative adequacy, as measured by the ratio of right minus wrong divided by total attempted ($t = 4.99$, $df = 30$, $p < .001$). This finding is even more impressive in view of the fact that the task seems to be graded in order of increasing difficulty. Thus the greater number of designs attempted by the young necessarily included proportionately more difficult items.

An effort was made to determine whether the types of error provide any further clues as to the differential efficiency of our samples on Concealed Figures. However, the data on hand allow only a very limited approach to this

TABLE 2
MEAN PERFORMANCES OF YOUNG AND OLD
GROUPS ON CONCEALED FIGURES

Group	N	At- tempted	Right	Wrong	Not Perceived When Present	Perceived When Not Present
Young	16	91.1	77.3	13.8	10.3	3.6
Old	16	55.5	36.9	18.6	13.9	4.7

question. A design is scored "wrong" under either of two conditions: 1. the simple figure is not perceived when present in the complex design, and 2. the simple figure is perceived as present in the complex design when such is not the case. For both young and old the first kind of error is decidedly more prominent. Moreover, the findings are similar for the two age groups; the proportion of each variety of error to total errors is approximately equal in both.

DISCUSSION

Perception is conceived to be a resultant of an interplay between field and individual factors. Concerned with demonstrating the importance of closure as one of the major organizing principles in perception, pioneer Gestalt psychologists (12, 13) employed configurations which maximized organizational forces *external* to the person; hence closure was compelling and instantaneous. In an incomplete but "pregnant" field, e.g., a circle with a small arc missing, the patterns themselves uniformly enjoined identical closure. Similarly, with a highly organized and stable field, as in Gottschaldt's original experiments (9), the existing closure was too coercive to segregate parts analytically. In these cases both effecting and resisting closure were viewed primarily as functions of field organization.

However, where field conditions are less coercive the manner and degree of closure is more likely to reflect the organizational skill of the observer. Consider first the problem posed by the Gestalt Completion Test. Here *S*'s task is essentially organizational: given a fragmented drawing he must consolidate the parts and visualize an integrated whole. Variations in the stimuli alone may make integration easier or more difficult, depending on the number, size and location of gaps, and the complexity of the entire drawing; but whether and how closure is achieved depends in large measure on the individual's ability. The situa-

tion in the Concealed Figures, which requires *S* to counteract tendencies toward closure, may be viewed similarly. To the extent that factors in the field enhance closures that are in conflict with a particular closure sought, the capacity of the individual to withstand and discard "erroneous" closure is revealed. Thus both the completion of and the resistance to closure are aspects of cognitive functioning which depend on *S*'s general integrative ability (16).

When dealing with Gestalt Completion, not only are the aged shown to be generally less adequate in developing closure, but, more specifically, their responses suggest a loose and/or overly rigid cognitive structure. Percepts are relatively vague and poorly integrated, bearing little resemblance to the figure delineated. They are also focused on details to the exclusion of other aspects of the drawing, perseverative and concretized in the sense that each segment given as a separate is treated as a separate and the relationships between adjoining segments are not envisaged.

Evidences of similar cognitive process in the aged were encountered by us on a task involving the judgment of ambiguous stimuli (14). The *Ss* were instructed to respond either "cat" or "dog" to each of a series of 13 line drawings graded through successive small modifications from a clear cat at one extreme to a clear dog at the other. Older persons, in the face of the progressively changing information provided within this order of cards, tended to exhibit a later point of shift in judgment as well as considerably more vacillation in sequence of judgments than younger individuals.

The deficit of aged persons in organizing ability is described perhaps even more strikingly in Rorschach investigations of senescence. Because of the greater ambiguity and lack of structure in the stimuli and the increased demand that organization be provided by *S* it is not surprising that repeated studies have found a distinct constriction and stereotypy of thought, confusion and cautiousness (4, 5, 12, 18).

Unlike the synthetic ability needed on Gestalt Completion, for Concealed Figures *S* must adopt an analytic attitude in order to discover whether the simple figure is embedded in the complex design. Each complex design contains many lesser figures which overlap and are

interwoven in intricate patterns. Identical lines and combinations of lines form one or more sides of a variety of different possible figures having the attribute of good form. The entire configuration itself is a harmonious whole. Thus the task basically taps the ability to maintain and recognize a given closure despite the intrinsic distractions of a more stable larger field and equally good competing gestalten. It is in this particular sense that the aged are found wanting in flexibility of closure.

A special problem arising is that of explaining the matter of omissions. The presence of omissions on Gestalt Completion is probably attributable to the instructions which, it will be recalled, permitted the option of leaving the item blank if it proved too difficult. But our two groups exercised their option in different measure. The old produced a significantly higher frequency of failures to respond than the young. We would hypothesize that this finding reflects one or both of the following characteristics in the aged: 1. greater impotence in the closure function, a phenomenon already witnessed in the analyses of responses actually given; and 2. a more cautious attitude, i.e., the need for a high degree of certainty before committing oneself. This tendency to inhibit response may represent a generalized adaptive mechanism through which the aged person avoids recognizing his inadequacy. Similar behavior in this same group was evidenced on a verbal learning experiment included in our battery (15). In paired-associate learning the older individuals, both initially and on repeated trials, much more usually gave "no response" when the stimulus item appeared.

However, unlike Gestalt Completion, there were no omissions on Concealed Figures. The reason is apparent in the set provided *S* by his assignment. He searches the large design to determine whether it incorporates the smaller figure. If he discovers that it does, he checks the item; if he cannot perceive that it does, then he assumes the figure does not exist within the large design and marks the item accordingly. Under these conditions a failure to respond is therefore not likely.

SUMMARY

A young and old group of 16 *Ss* each, mean ages 26.8 and 78.1 years and equated for intel-

ligence, were compared in performance on two tests of perceptual closure, Gestalt Completion and the Thurstone adaptation of Gottschaldt's Concealed Figures. The former procedure requires the capacity to organize, the latter to resist closure. Older *Ss* were distinctly inferior on both tasks. Moreover, an analysis of "wrong" responses on Gestalt Completion revealed signs of perseveration, fixation on details and concretization by the aged individuals, as well as a more general tendency toward vagueness and poor articulation of form. The old also omitted a significantly greater number of items on Gestalt Completion. This finding was interpreted as indicating one or both of the following: further evidence of a lessened integrative ability, and excessive cautiousness, i.e., a defensive reluctance to venture response for fear of recognizing their inadequacy. The over-all results were viewed as reflecting the overly loose or overly rigid cognitive functioning of the aged.

REFERENCES

1. BABCOCK, HARRIET, & LEVY, L. *Test and manual of directions, the revised examination for the measurement of the efficiency of mental functioning*. Chicago: C. H. Stoelting, 1940.
2. BOTWINICK, J. & BIRREN, J. E. Differential decline in the Wechsler-Bellevue subtests in the senile psychoses. *J. Geront.*, 1951, **6**, 365-368.
3. CALDWELL, BETTYE M. The use of the Rorschach in personality research with the aged. *J. Geront.*, 1954, **9**, 316-323.
4. CHESROW, E. J., WOSIKA, P. H., & REINITZ, A. H. A psychometric evaluation of aged white males. *Geriatrics*, 1949, **4**, 169-177.
5. DAVIDSON, HELEN H., & KRUGLOV, LORRAINE. Personality characteristics of the institutionalized aged. *J. consult. Psychol.*, 1952, **16**, 5-12.
6. DENNIS, W. Age and behavior. *USAF Sch. Aviat. Med. Proj. Rep.*, 1953, Proj. No. 21-0202-0005 (Rep. No. 1).
7. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
8. GILBERT, JEANNE G. *Understanding old age*. New York: Ronald, 1952.
9. GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. I. *Psychol. Forsch.*, 1926, **8**, 261-317.
10. HOWELL, R. J. Changes in Wechsler subtest scores with age. *J. consult. Psychol.*, 1955, **19**, 47-50.
11. KLOPFER, W. Personality patterns in old age. *Rorschach Res. Exch.*, 1946, **10**, 145-166.
12. KOFFKA, K. *Principles of gestalt psychology*. New York: Harcourt, Brace, 1935.
13. KÖHLER, W. *Gestalt psychology*. New York: Liveright, 1947.
14. KORCHIN, S. J., & BASOWITZ, H. The judgment of

- ambiguous stimuli as an index of cognitive functioning in aging. *J. Pers.*, 1956, **25**, 81-95.
15. KORCHIN, S. J., & BASOWITZ, H. Age differences in verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 64-69.
16. KORCHIN, S. J., SINGER, J. L., & BALLARD, R. G. The influence of frustration on the reproduction of visually-perceived forms. *Personality*, 1951, **1**, 54-66.
17. MILES, W. R., & MILES, CATHARINE C. Principal mental changes with normal aging. In E. J. Steiglitz (Ed.), *Geriatric medicine*. Philadelphia: Saunders, 1943. Pp. 99-117.
18. PRADOS, M., & FRIED, E. Personality structure in the older age groups. *J. clin. Psychol.*, 1947, **3**, 113-120.
19. STREET, R. F. *A gestalt completion test*. New York: Teachers Coll., Columbia Univer., 1931.
20. THURSTONE, L. L. *A factorial study of perception*. Chicago: Univer. of Chicago Press, 1944.
21. THURSTONE, L. L. *Mechanical aptitude III. Analysis of group tests*. Chicago: Univer. of Chicago, 1949. (*Psychometr. Lab. Rep.* No. 5.)
22. WECHSLER, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Received March 19, 1956

POSTDECISION EXPOSURE TO RELEVANT INFORMATION¹

DANUTA EHRLICH, ISAIAH GUTTMAN, PETER SCHÖNBACH,

University of Minnesota

AND JUDSON MILLS

Stanford University

UNDER what conditions will persons voluntarily expose themselves to information? This is an important theoretical and practical question for those interested in social influence processes. It is widely accepted that the audiences of mass media are to a large extent self-selected. However, there has been little rigorous specification or systematic research concerning the variables related to seeking out or avoidance of new information. Festinger has recently proposed a theory (2) which makes some definite predictions about selective exposure to information following decisions, and it is the purpose of this study to test several of them. Before the specific predictions from the theory are stated the relevant parts of Festinger's theory will be summarized briefly.

The theory is concerned with relations among cognitive elements, that is, things which a person knows about himself or his environment. For example, knowing that one is on a picnic is a cognition about one's behavior; knowing that it is raining, a cognition about the environment. Cognitive elements may be either consonant, dissonant, or irrelevant to one another. A behavioral element is consonant with an environmental element if, considering these elements in isolation, the behavioral would follow from the environmental element; dissonant if, considering them alone, it would *not* follow. Thus knowing that it is raining is ordinarily dissonant with going on a picnic; consonant with staying home.

According to the theory, the presence of dissonance gives rise to pressure to reduce or eliminate it. The strength of these pressures is a function of the magnitude of the dissonance. The theory also states that the greater the importance or the value of the elements in-

involved, the greater is the magnitude of the dissonance and hence the greater the pressure to reduce it.

The total magnitude of the dissonance between a behavioral element and the cluster of relevant environmental elements, assuming they are all equally important, is equal to the proportion of dissonant elements in this cluster. Therefore, one way that a person can reduce dissonance is to add new environmental elements that are consonant with his behavior. Thus the theory predicts that persons with dissonance in general seek out consonant information and avoid information that would introduce new dissonant elements.

When a person makes a decision, a corresponding behavioral element is established and his cognitions about the alternatives among which he has chosen are then consonant or dissonant with this element. All the favorable aspects of the unchosen alternatives and unfavorable aspects of the one chosen are dissonant with the choice. Consequently, the creation of dissonance is a common result of decisions. After the choice has been made, the person therefore tends, according to the theory, to expose himself to information that he perceives as likely to support the decision and to avoid information that is likely to favor the unchosen alternatives.

The purchase of a new automobile, for example, is usually a rather important decision for a person. Considerable dissonance should exist for a new car owner immediately after he has bought his car; all "good" features of the makes he considered, but did not buy, and "bad" features of the one he bought are now dissonant with his ownership of the car. He should also attempt to reduce this dissonance. In this instance of postdecision dissonance, sources of information in support of the decision are readily available. Since automobile advertising contains only material favoring the particular car advertised, reading advertisements of his own make is one way a new car owner can get information supporting his

¹ This research was carried out under the direction of Dr. Leon Festinger while the authors were members of his graduate seminar at the University of Minnesota. They would like to thank him for his generous advice and constructive criticism. This study was also part of the research program of the Laboratory for Research in Social Relations.

choice and thereby reduce his dissonance. On the other hand, reading advertisements of other cars may increase his dissonance.

The specific hypotheses of the study were as follows:

1. After a decision persons tend to seek out dissonance-reducing information. Thus new car owners will read advertisements of their own cars more often than those of (a) cars which they considered but did not buy and (b) other cars not involved in the choice.

2. After a decision persons tend to avoid dissonance-increasing information. Thus new car owners will read advertisements of considered cars less often than of other cars.

3. Postdecision dissonance is, in general, reduced over a period of time. Thus these selective tendencies in readership of advertising will not exist for owners of old cars.

The theory predicts other consequences of postdecision dissonance, for example, changes in desirability ratings of chosen and unchosen alternatives as investigated by Brehm (1). Our present concern is limited to the effect of decision-making on exposure to information.

METHOD

One hundred and twenty-five male residents of Minneapolis and its suburbs, who owned one of eight popular automobile makes, were interviewed² in an "advertising survey," after an appointment had been made by telephone. Sixty-five of the respondents were new car owners chosen randomly from a list of recent auto registrations, dated four to six weeks before the time of the interview. The other 60 respondents owned cars manufactured in 1952 (three years before the study was done) or before, and were selected from the Minneapolis telephone directory. (Actually about a third of the old car owners bought their cars after 1952; however, none of the purchases were made during the year in which the study was done.) The number of owners of different makes of cars was approximately equal in both groups.

During the course of the interview:

1. Each respondent was asked to recall recent automobile advertisements which had impressed him particularly and to say where and when he had seen the advertisement, whether or not his impression was favorable, and how much he had read (Recall Data).

2. Each was shown issues of popular magazines which appeared four weeks or less before the interview (after the new car owners had made their decisions) and copies of one of the two Minneapolis newspapers for seven days prior to the interview. Only those publications were included which the respondent previously indicated as having read regularly. For each issue the respondent was asked if he had seen it, and

if so, he was shown each car advertisement it contained, asked whether he had noticed the advertisement, and, if he had, whether he had "read all," "read some," or "just glanced at it" (Recognition Data).

3. Each was presented with eight large plain envelopes bearing the names of the eight popular makes. The respondent was told that his reactions to some "new automobile advertising" were desired and asked to choose two of the envelopes in order to read and comment upon the material they contained (Envelope Data).

At the conclusion of the interview the respondent was asked to name other makes which he "seriously considered" before he decided to buy his present car.

RESULTS

As a prerequisite for the predicted selectivity in a person's exposure to information, he must have some expectation about the likelihood that a given item will reduce or increase dissonance for him. Thus the new car owners must first recognize that a particular make is being advertised, i.e., they must at least notice the advertisement. For each respondent the percentage of advertisements which he read, of those which he noticed, was therefore calculated separately for his own car, the cars he considered, other cars not involved in the choice, and also for those considered and other cars combined. (The categories "read all" and "read some" were combined because there were two few advertisements reported as read completely to make a separate analysis feasible; results are essentially the same for these categories.) The means of these percentages for the two groups and the corresponding *N*'s are reported in Table 1, which also contains

TABLE 1
MEAN PERCENTAGES OF ADVERTISEMENTS
NOTICED AND OF ADVERTISEMENTS
READ OF THOSE NOTICED
(Corresponding *N*'s appear in parentheses)

(Corresponding N's appear in parentheses)				
Car Owners	Make Advertised			Considered Plus Other
	Own	Considered	Other	
Percentage Noticed				
New	70 (52)	66 (50)	46 (64)	48 (64)
Old	66 (51)	52 (31)	40 (60)	41 (60)
Percentage Read of Those Noticed				
New	67 (47)	39 (46)	34 (64)	35 (64)
Old	41 (44)	45 (20)	27 (57)	30 (57)

Note.—The *N*'s are reduced because in some cases no advertisements of a particular kind appeared in the issues shown or none of those which appeared were noticed. They are further reduced because not all respondents named cars as "seriously considered."

² We would like to thank Craig Erickson and Peter Liebes for their assistance in conducting the interviews.

TABLE 2

SIGN TEST COMPARISONS BETWEEN PERCENTAGE OF ADVERTISEMENTS READ OF DIFFERENT CATEGORIES

Comparison	Car Owners		Comparison	Car Owners	
	New	Old		New	Old
% Own Minus % Considered	+ 21 - 4 $p < .01$	5 6 n.s.	% Own Minus % (Considered Plus Other)	+ 32 - 7 $p < .01$	20 17 n.s.
Between Groups $p < .05$ ($\chi^2 = 3.90$)*			Between Groups $p < .01$ ($\chi^2 = 6.89$)		
% Own Minus % Other	+ 31 - 8 $p < .01$	20 16 n.s.	% Considered Minus % Other	+ 19 - 15 n.s.	10 7 n.s.
Between Groups $p < .03$ ($\chi^2 = 4.93$)			Between Groups, n.s.		

* Corrected for continuity.

means for the percentage of advertisements noticed of those which were shown.

It can be seen from Table 1 that new car owners read advertisements of their own make more often than advertisements of both considered and other cars. These differences do not appear among old car owners; in fact the difference between reading of own and considered car advertisements is in the opposite direction. However, it is obvious from Table 1 that new car owners do not read advertisements of considered cars less often than of other cars. Finally, it is apparent that readership of advertisements, in general, is greater among new car owners.

The significance of these differences was determined by means of the sign test. For each respondent the sign of the comparison between the percentage of advertisements read of the different categories was found and this distribution of signs was tested using the critical values given in Mosteller and Bush (3). Results of these comparisons appear in Table 2 which shows that the differences apparent from Table 1 are significant. New car owners read significantly more advertisements of their own car than of both considered and other cars; none of the other comparisons yield significant differences. It can be seen that the N 's have been further reduced because these comparisons were not possible for all respondents.

The significance of the differences between new car and old car owners was tested by means of chi square. Results of these tests are also given in Table 2. The comparison *own vs. considered* is in the predicted direction with $p < .05$. The comparison *own vs. other* is significant at the .03 level, and the comparison *own vs. considered plus other* is significant at the .01

level; there is no significant difference for the comparison *considered vs. other*.

One other comparison in the recognition data is worth noting.³ New car owners were divided into two groups; those who named less than two cars as "seriously considered" and those who named two or more. The difference in reading *own vs. considered plus other* was examined separately for these two groups and was found to be greater for those who considered more than two cars. That is, new car owners who named more than one considered car read relatively more advertisements of their own car. For this group the mean difference of the percentages was 37; that of the other group was 26. The difference between the two groups was tested by a median test and the resulting (two-sided) $p = .19$ (12 of the 19 who considered two or more were above the median; only 12 of the 28 who considered less than two were above it).

For the Envelope Data the difference between groups is in the predicted direction but not significant. Fifty-one of 60 (85%) new car owners chose an envelope of their own car, 45 of 59 (76%) old car owners chose one of their own car. Respondents in both groups obviously chose to read and comment upon "new advertising" of their own makes far more often than would be expected by chance.

The number of impressive advertisements (Recall Data) which the respondents read is too small to make analysis of these data possible. (Of course no advertisements were included which appeared before the new car

³ See the discussion section for the theoretical relevance of this comparison.

owners made their decisions or before a comparable length of time for old car owners.)

DISCUSSION

The data from the recognition of advertisements which appeared in recent publications (after the new car owners made their decisions) strongly support the prediction of Festinger's theory that, after an important decision, persons seek out consonant information. New car owners were found to read advertisements of their own car more often than those of cars they considered but did not buy, and of other cars not involved in the choice. That this selective tendency in exposure to information is not simply a result of car ownership but is related to decision making is demonstrated by its failure to occur among old car owners. The Envelope Data also indicate that new car owners prefer to read advertisements of their own car although the difference between new and old car owners is not significant.

An alternative explanation of the results might be offered. If people do not remember accurately when they read an advertisement, and if shortly *before* buying a car, persons are more likely to read advertisements of *that* car than of others, then the advertisements of their own make may have actually been read before the decision and the obtained differences may be a result of the new car owners' inability to remember the issue in which they read them. This interpretation seems unlikely since it is plausible to suppose that, *before* their purchase, the new car owners attempted to learn as much as possible about all the different makes involved in their choice and therefore read advertisements of the makes that they considered as often as those of the make that they finally chose. Nevertheless, the possibility that it could account for the data was checked by determining whether the advertisements of the make actually chosen by the new car owners which they said they read also appeared during the month prior to their purchases. Only four of 54 such advertisements also appeared in this period. When these few instances in which the alternate explanation might apply are eliminated from the analysis the results are essentially unchanged.⁴

⁴ There are only two small changes in the significance tests reported in Table 2. When the four ads are eliminated the χ^2 s for the comparisons between

The prediction that new car owners would read fewer advertisements of considered cars than other cars was not substantiated. However, the data are not sufficient to reject conclusively the hypothesis that dissonance-increasing information is avoided after a decision. It is possible that some of the new car owners expected advertisements of considered cars to decrease—not increase—their dissonance. They may have read the advertisements of the rejected cars not with the intention of getting further information about their advantages but for the purpose of finding fault with them, comparing them unfavorably with the chosen car. For example, an owner of a make with high horsepower who regards horsepower as an important feature of automobiles may read advertisements of considered cars of lesser horsepower to remind himself of their inferiority in this respect. We would not expect this to occur frequently but it might happen often enough to account for the failure of the prediction. The data do not provide any test of this interpretation.

The indication that new car owners who named more than one considered car tend to read relatively more advertisements of their own car than do those who named one or none is additional, suggestive evidence for Festinger's theory. Since the total magnitude of dissonance is a function of the proportion of dissonant elements, we would expect that these persons would in general have greater dissonance, for as the number of unchosen alternatives increases the proportion of dissonant elements increases. Therefore the theory predicts that they should show a greater tendency to seek out consonant information.

SUMMARY

Readership of auto advertising by new and old car owners was investigated in order to test some predictions of Festinger's theory of dissonance concerning selective exposure to information following decisions. It was found that new car owners read advertisements of their own car more often than of cars they considered but did not buy and other cars not involved in the choice. These selective tenden-

groups for *own minus considered* and for *own minus considered plus other* are reduced to 3.61 and 5.90 respectively.

cies in readership were much less pronounced among old car owners. This finding supports the theoretical derivation that persons in general seek out consonant or supporting information after an important decision in an attempt to reduce dissonance resulting from it.

From the derivation that persons tend to avoid dissonance-increasing information it was predicted that new car owners would read advertisements of considered cars less often than of other cars. The data do not confirm this prediction. It was suggested that some of the new car owners may have expected the

considered car advertisements to decrease—not increase—their dissonance.

REFERENCES

1. BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, **52**, 384-389.
2. FESTINGER, L. The relation between cognition and action. Paper read at Symposium on Cognition, Boulder, Colo., May, 1955.
3. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. I. Cambridge: Addison Wesley, 1954.

Received March 19, 1956.

A NEW LOOK AT VIGILANCE AND DEFENSE¹

DONALD P. SPENCE²

Teachers College, Columbia University

As New Look experiments continue to accumulate, the exact nature of perceptual vigilance and defense remains open to question. Various investigators (1, 2, 7) have shown that anxiety-provoking stimuli are recognized at both higher and lower thresholds than control stimuli, but the meaning of these threshold deviations is still unclear and their reliability not established. Findings are ambiguous partly because (a) the affective qualities of the so-called "taboo" and control stimuli are usually assumed, not predetermined (4, 5, 8); (b) the frequency counts generally used often do not apply specifically to the experimental sample; and (c) thresholds are usually measured only once, after the experimental threat has been introduced (5, 8). There is thus no check on whether the stimuli were truly matched, and thus no means of partialling out the effect of residual structural differences in "taboo" and control words.

Apart from these procedural shortcomings, the question still remains as to what processes contribute to vigilant and defensive responses. Some studies (1, 2) have succeeded in showing that threshold scores are correlated with personality factors, but the basis for the correlations has not yet been adequately explained. Vigilance and defense have not been explicitly distinguished in terms of the perceptual operations involved, and the possibility exists that they are two manifestations of the same underlying variable.

As a start toward a single-factor theory, it might be argued that both vigilance and de-

fense result from the disrupting effect of the threatening stimuli. Anxiety associated with the "taboo" word, for example, interferes in some way with the S's normal hierarchy of responses to a partially exposed stimulus. This disruption has the effect of reshuffling the usual order of prerecognition hypotheses. As a result, the correct response is given much earlier (low threshold) or much later (high threshold) than it would occur for the matched control word. By this reasoning, anxiety is related to both threshold elevation and depression, and could account for both vigilant and defensive reactions.

These assumptions suggest that words known to create anxiety are recognized at either higher or lower thresholds than control words. The anxiety associated with the threatening stimuli is related to the absolute amount of threshold deviation, but not to either vigilance or defense separately. When the associated anxiety is removed, threshold deviations are expected to diminish.

The present experiment was designed to test these assumptions while taking into account the major sources of error mentioned above. Threat and control words were carefully equated with respect to length, frequency, and ease of recognition before experimental failure was introduced. The experimental anxiety was systematically associated with the threat words, and its effect independently measured. The following hypotheses were tested:

I. Threatening tachistoscopic stimuli are recognized at thresholds *both* higher and lower than thresholds for matched control words. The extent of deviation is related to the degree of anxiety induced by the threat.

II. When the associated anxiety is removed, then threshold deviations diminish. The degree of reduction is related to the decrease in anxiety.

METHOD

Twenty-two adults, five men and seventeen women were tested in individual sessions averaging two and one-half hours apiece. Each session was divided into three approximately equal parts, separated by two ten-minute intermissions. All Ss were paid for their

¹ This paper is based on a dissertation submitted in candidacy for the degree of Doctor of Philosophy at Teachers College, Columbia University. The writer is indebted to members of his thesis committee, Helen M. Walker, Paul E. Eiserer, and especially E. J. Shoben, Jr., Chairman, for their suggestions and encouragement; to Gardner Murphy who made this study possible by extending his advice, criticism, and the research facilities of the Menninger Foundation; and to David T. Herman and Joseph E. Brewer for making possible initial pilot work at the University of Wichita. This study was done while the writer was a research assistant at the Menninger Foundation.

² Now with the Research Center for Mental Health, New York University.

TABLE 1
STIMULUS WORDS

abbot*	hinge	slant*
acorn*	hoist*	snare
baker	kites*	snowy*
beech*	liver	stake
braid	niece*	stung*
click*	orbit*	tarry*
flake*	piper*	tutor
grope*	posed	usher*
handy*	prose	

* Potential failure words which cannot be meaningfully rearranged. Each S's set of failure words was chosen from this group.

time. The majority were persons associated in some way with the Menninger Foundation where the research was conducted—friends and wives of staff; aides; nurses; and outpatients. Five girls in secretarial training completed the group. Whether or not an S was an outpatient did not appear to affect subsequent results, so outpatients were included in the final sample.

Twenty-six five-letter words of relatively neutral content and equivalent Thorndike-Lorge frequency were used as stimuli (see Table 1). Sixteen of these words could not be rearranged to form other words (e.g., CLICK) and they constituted the potential failure list. All words were presented initially to the S in a random sequence to determine base recognition threshold (Trial I). Words from the failure list (insoluble) were then matched with words of equal threshold from those remaining and all unmatched words were discarded. The matching was done individually for each S, providing him with a tailored set of six to eight pairs of words, depending on the number of failure words recognized initially.

Recognition thresholds were determined by gradually increasing the brilliance of a word as it was being exposed at $\frac{1}{100}$ of a second. Change in illumination was controlled by a Variac in parallel with a 75-watt bulb. A Triplett AC volt-meter served as a check on random fluctuations of line voltage. Words were initially exposed with the Variac set at 30 volts and gradually increased two volts per exposure until recognition. Stimuli were typed in capitals and photographed; the resulting black and white negative was fitted into a slide holder and presented 20 inches from the S. It was illuminated from behind and mounted in a long oblong box, $28 \times 3\frac{1}{2} \times 3\frac{3}{4}$ inches, open at one end and painted flat black on the inside. Covering the slide was a strip of ground glass to minimize early recognition. Two strips of fluorescent paint outlined the stimulus to orient the S. Thresholds were measured in absolute darkness; two five-letter practice words were exposed first to familiarize S with the procedure. He was then told to expect 26 additional words of similar length. Each word was presented individually at successive levels of illumination until it was correctly recognized. One correct response was considered sufficient for establishing a threshold.

After all words had been recognized and paired according to their thresholds, each control word was presented to S on a white card for 30 seconds. He was instructed to examine the word and mentally rearrange its letters (e.g., DIMLY:MIDLY). He was not directed

to form a real word because (a) not all control words could be meaningfully rearranged; and (b) earlier pilot work indicated that failure to rearrange a word was unsettling, even if S were told that it was a practice session. Therefore, S was instructed to rearrange each word in whatever way he preferred, as follows:

This is an exercise in visualizing words. There are no right or wrong answers. It is merely a warm-up session to get you accustomed to thinking of words in different arrangements. The idea is to look at a word and to see in what ways you can rearrange the letters. Obviously, there are a large number of possible combinations you might think of, but I don't expect you to exhaust them all. I merely want you to imagine several arrangements for each word. For example, take the word DIMLY. You might think of it with the D and the M interchanged to make MIDLY. Some arrangements will make sense and others will not; the end result doesn't matter. Take each word as I give it to you and practice rearranging the letters. You don't have to announce each new combination as you find it; make a mental note of the solution and try another. You will have 30 seconds for each word.

The DIMLY:MIDLY example was intentionally presented to form a set which would favor nonsense solutions. This precaution was introduced to discourage Ss from making meaningful arrangements from the potentially soluble words which might lead to differences in their later recognition. No S mentioned a meaningful solution and the chances are they would have if they had found one because the instructions, by implication, suggest there are few. Moreover, no S showed signs of boredom or frustration so it may be assumed that the control task was, as intended, essentially neutral.

Another, more difficult, practice task was presented following the control series. Additional words, chosen for their solubility, were given the S with instructions to make a new word from each (e.g., SPINE:PINES). After the S appeared quite confident of his ability to form anagrams, he was given the failure words with some soluble words added to increase the realism of the task. He was given these instructions:

All right, your performance on the practice series was excellent. Now I am going to give you a test list. Judging by your performance on the practice list you just finished, you should be able to form one new word from each of the words on the test list. Tell me each word as you find it. You will have 30 seconds for each one.

As the S failed each of the failure words, the E would sharpen this feeling by asking "Why aren't you doing better? You did well on the practice task," and other, similar questions. Most Ss became increasingly disturbed by the task and occasionally offered nonsense solutions when they could think of nothing else.

For the majority of Ss the failure task achieved its intended, depressing effect and its results persevered for the remainder of the experiment (see below). In all other respects the failure and control tasks were essentially similar. In each case S was presented with a series of words, each of which he attempted to rearrange within a 30-second period. Words were matched for

length, frequency, and recognition threshold. The *S* was not asked to verbalize his solutions on the control task; he could not do so on the failure task. He tended to form nonsense anagrams on the control task (because of the instructions) and was forced to give similar responses to the failure task (because no other solution was possible). Failure and control experiences were essentially different only with respect to failure, and it seems reasonable to explain subsequent experimental differences in terms of this variable.

Immediately after the failure, each *S* was presented with an Adjective Check List (ACL) of 145 words adopted from Nowlis *et al.* (6). Included in the list were two groups of words identified by Nowlis on an empirical basis: an Anxiety cluster and an Elation cluster of 10 words each. The *S* was asked to check the words which best described his mood. He was then questioned about his feelings, and after a short interpolated task ("List all the words you can think of beginning with Q"), he was asked to list all words he could remember from the failure and control lists.

After a ten-minute rest period during which the experimenter left the room, *Ss'* recognition thresholds for the failure and control lists were measured a second time (Trial II). Words were exposed in random order, each to a criterion of one correct recognition. The *S* was then informed that all the anagrams he had failed were actually impossible and that he had been forced to fail because the experimenter was interested in the perceptual consequences of failure. He was encouraged to speak about the experience at some length. When the *S* had expressed his feelings he was given the ACL a second time, and after another interpolated task, memory for failure and control words was again measured. Recognition thresholds were then measured a third time (Trial III) after another ten-minute rest.

RESULTS AND DISCUSSION

Before either hypothesis can be tested, it is necessary to establish that soluble and insoluble control words can be combined for purposes of analysis. Anecdotal reasons for assuming they could be treated together were presented earlier and as a further check, thresholds for soluble and insoluble words were compared on the assumption that if there were differences in the anagram experience they would appear in the recognition scores. Average thresholds after failure for soluble and insoluble control words were compared for each *S*; the difference for the group is not significant ($t = 1.46$; $p > .10$). Meaningful solutions to control words did not seem to inhibit later recognition. Moreover, average failure thresholds were not significantly different from either average thresholds for soluble control words ($t = .68$; $p > .20$) or average thresholds for insoluble control words ($t = 1.58$; $p > .10$). Accordingly, for each *S*, all thresholds for control words were combined and his average

control threshold was subtracted from his average threshold for failure words. A second difference score was computed from thresholds for failure and control words after the explanation. A positive score indicates that failure words were recognized, on the average, at higher levels than control words; a negative score indicates the reverse relationship.

Anxiety created by the failure was estimated by counting the number of words checked on the Anxiety cluster of the ACL after failure and dividing by the total number of adjectives checked (Anxiety Ratio I). Anxiety was similarly estimated from the second ACL after the explanation (Anxiety Ratio II). Elation created by the explanation was estimated by counting the number of words checked on the Elation cluster after the explanation and dividing by the total number of words checked.

Memory was scored by subtracting the number of control words from the number of failure words recalled after the failure; a similar score was obtained for recall after the explanation.

In accordance with the first hypothesis, anxiety expressed after the failure (Anxiety Ratio I) was significantly correlated with absolute differences between thresholds for failure and control words after the failure ($\rho = .57$, $p < .01$; see Table 2). Anxiety was uncorrelated with the tendency to recognize failure words

TABLE 2
RANK-ORDER CORRELATIONS BETWEEN
MAJOR VARIABLES

	Rho
Anxiety Ratio after failure vs. Threshold difference after failure (absolute value)	.57**
Threshold difference after failure (with sign)	.02
Threshold difference after explanation (absolute value)	.52**
Failure word consistency	.38*
Control word consistency	-.22
Memory after failure	.49*
Memory after failure vs. Threshold difference after failure (absolute value)	.50**
Threshold difference after failure (with sign)	-.10
Elation Ratio after explanation vs. Threshold difference after explanation (absolute value)	.29
Anxiety change vs. Threshold change	.11

* $p < .05$.

** $p < .01$.

For 22 cases, a rho of .359 is significant at the 5% level, and a rho of .508 is significant at the 1% level.

above control words ($\rho = .02$); hence, anxiety and "perceptual defense" seem unrelated. On the contrary, *Ss* with high anxiety scores recognized control words at both higher and lower thresholds than failure words, with the absolute degree of difference a function of their anxiety.³ Table 3 indicates how absolute threshold differences increase as a function of anxiety. Since the failure and control words were equated with respect to frequency, length, and ease of recognition, the differences in perception after failure cannot be attributed to these residual factors. Moreover, suppression of verbal report cannot account for the findings as suggested by Whittaker, Gilchrist, and Fischer (11) because the significant correlation is with recognition deviation, not with recognition delay.

When thresholds alone are considered, recognition of failure and control words seems to be randomly determined. Thirteen *Ss* recognized failure words at average thresholds below control words ("vigilants"); eight *Ss* showed the reverse, or "defense" reaction; and one *S* did not discriminate between lists. For all *Ss*, average failure-word thresholds are not significantly different from average control-word thresholds; $p > .3$; two-tailed sign test (10). The present dichotomy of "vigilant" and "defensive" *Ss* parallels the findings of Postman and Solomon (7), Chodorkoff (2), and others, and supports the general notion that unpleasant stimuli are not necessarily perceived at elevated thresholds. Such data are usually explained by assuming that some *Ss* are "vigilant" and others "defensive." The present findings call this interpretation into question because of the mixed nature of individual responses. No *S* was exclusively "vigilant" or "defensive" in his recognition of all words; rather, each *S* recognized some failure words above, and some below, matched control words.⁴ The random distribution of differ-

³ It might be argued that more anxious *Ss*, because they seem more sensitive to the differences between failure and control words, would also differentiate between soluble and insoluble control words. They do not. Anxiety after failure is not significantly correlated with the absolute difference between thresholds for soluble and insoluble control words.

⁴ The tendency toward recognizing all failure words above or below matched control words correlated $-.01$ with Anxiety Ratio I. Thus, whether an *S* exhibited a preponderance of "vigilant" or "defensive" responses is not related to anxiety.

TABLE 3
MEAN FAILURE AND EXPLANATION ABSOLUTE
THRESHOLD DIFFERENCE SCORES

Group*	Absolute Threshold Difference Score After Failure** (Mean)	Absolute Threshold Difference Score After Explanation*** (Mean)
High Anxiety	3.39	3.12
Medium Anxiety	1.68	1.16
Low Anxiety	1.20	.73

* Group placement determined by score on Anxiety Ratio I.

** Between groups, $\chi^2 = 6.6$, $p < .05$.

*** Between groups, $\chi^2 = 9.0$, $p < .02$.

Note.—Significance tests based on an analysis of variance for ranks described in Walker (10).

ences for all *Ss* makes it difficult to argue that a "vigilant" or "defensive" average response is characteristic of a particular personality style. The tendency to recognize failure words above or below control words appears to be dependent on factors other than the particular characteristics of the *S*.

Absolute size of threshold differences appears to be also important in relation to recall. The tendency to remember more control than failure words after the failure correlated $.50$ with the absolute threshold difference ($p < .05$). When direction of difference was taken into account, the correlation dropped to $-.10$. Thus an *S* who forgot relatively more failure words did not necessarily perceive failure words at generally higher levels than control words. Forgetting of failure words was, however, directly related to anxiety. The *Ss* with high anxiety scores tended to recall relatively more control words after the failure ($\rho = .49$; $p < .05$). These results are strongly suggestive of findings by Glixman (3) who tested memory for failure under three degrees of stress. He found that the number of incomplete (failure) tasks recalled varied inversely with the severity of the stress situation; i.e., with the extent to which the instructions stressed the personal implications of the task. Present results bear out his findings to the degree that forgetting of failure (stressful) words is a function of anxiety.

Did threshold differences disappear when anxiety was reduced, as predicted by Hypothesis II? Sixteen *Ss* had smaller anxiety scores on the ACL after the explanation (Anxiety Ratio II) than on the ACL after the failure (Anxiety Ratio I), but only ten of these

had smaller average threshold differences after the explanation. Change in anxiety scores correlated .11 with change in threshold differences. The elation scores correlated only .29 with threshold differences after the explanation. It is possible that the explanation did not produce relief of a sufficient degree to test the hypothesis properly, in which case anxiety created by the failure should still be evident after the explanation. A significant correlation ($\rho = .52$; $p < .01$) between failure-induced anxiety (Anxiety Ratio I) and threshold differences after the explanation suggests that such is the case. Enough anxiety was left unaffected by the explanation to influence thresholds in Trial III (see Table 3). The presence of lingering anxiety is understandable in view of the impact of the failure task and the brief time allotted to explanation.

The high correlation with initial anxiety suggests that because of the cumulative effects of the initial failure, the experiment did not provide a satisfactory test of Hypothesis II. Threshold and memory scores after the explanation are probably some function of both the explanation and the earlier failure, and because of the difficulty in their interpretation, the remainder of the paper will be devoted to an analysis of threshold changes after failure. The high correlation, furthermore, is surprising since one would expect scores on repeated trials to approach some asymptotic limit and argue, in consequence, that since thresholds for control and failure words were originally equated in Trial I, they should again approach equality after repeated testing. That they did not is further evidence of the strength of the experimental failure.

If threshold differences after the failure are related to anxiety, it can be argued that these differences reflect the impact of anxiety upon the perception of the failure words. Accordingly, thresholds for failure words should change more than thresholds for control words from Trial I to Trial II, with change a direct function of anxiety. Further analysis does not support this assumption. Anxiety is not related to change in failure word thresholds, but rather to lack of change, or consistency of recognition. Contrary to expectations, Ss with high scores on Anxiety Ratio I recognize failure words after failure at thresholds which correlate highly with thresholds before the failure.

The higher the anxiety, furthermore, the higher the consistency, and the rank-order correlation between intertrial consistency of failure word thresholds and Anxiety Ratio I is .38 ($p < .05$). Comparable intertrial consistency of control word thresholds, before and after the failure, is not related to Anxiety Ratio I ($\rho = -.20$). Thus it would appear that anxiety is inversely related to failure-word variability and unrelated to control-word variability, suggesting that control-word thresholds can be interpreted as a recognition baseline for all Ss, unconfounded by the effects of anxiety.⁵

Apparently, then, the anxious S after the failure did not substantially alter his choice or sequence of hypotheses in recognition of the failure words. Anxiety stabilized rather than disrupted perception. Although surprising in view of the assumption mentioned earlier that anxiety is disruptive, this finding is in line with recent studies of anxiety and habit strength. Taylor and Spence (9) have suggested that anxiety may reinforce the habit strength of both correct and incorrect responses. Applied to the present study, their findings help explain why Ss who are anxious following the failure should respond to the failure words essentially as they did initially. Apparently the failure reinforced both correct and incorrect guesses for failure words with the result that recognition thresholds after the failure correlated highly with thresholds before failure. The Taylor and Spence findings might also explain the low intertrial correlation of failure-word thresholds for unanxious Ss. They presumably responded more flexibly to the failure; some, but not all of their initial responses were reinforced; consequently, their thresholds after the failure were uncorrelated with earlier thresholds.

This line of reasoning needs to be checked by future investigation, but one thing seems clear. The anxiety associated with threatening words apparently leads to an inhibition of

⁵ Intertrial consistency of control word thresholds from before to after the failure is not significantly different from intertrial consistency of failure word thresholds over the same period ($p > .3$, two-tailed sign test [10]) which argues against the possibility that recognition of control words after the failure was hindered by associative linkages formed during the control task. Each S might have formed more solutions to control words than to failure words during the anagram tasks, but this differential productivity did not seem to affect subsequent recognition.

response; this inhibition may result in either "vigilance" or "defense." What are usually assumed to be two distinctive reactions to tachistoscopic stimuli seem, on closer examination, to be two varieties of the same underlying process. The further fact that no *S* was exclusively "vigilant" or "defensive" in his reaction to the failure words makes it difficult to argue that these ideal response types can be used to discriminate between individuals. If there are distinctive individual ways of reacting to anxiety-provoking stimuli, they seem partly related to the amount of anxiety aroused in the *S*. Further experiments are needed using other measures of anxiety to establish these findings more securely.

SUMMARY

Two lists of words matched for frequency, length, and recognition threshold were presented in an anagram-solving exercise. After one list had been failed, all words were exposed tachistoscopically for a second time. Anxiety aroused by the failure correlated significantly with the absolute difference between recognition thresholds for failure and control words. Anxiety did not correlate with either "vigilant" or "defensive" responses alone, nor were failure words recognized at significantly different levels from control words when the anxiety score was not considered. The relationship between anxiety and absolute threshold difference persisted even after the artificial nature of the failure had been explained, contrary to the hypothesis that explanation would erase the effects of failure.

Sequential analysis of thresholds indicated that successive recognition of failure and control words did not follow a simple pattern. Anxiety appeared to stabilize the repeated

recognition of failure words, with threshold consistency a function of anxiety. It can be argued, after Taylor and Spence (9), that anxiety tends to reinforce the use of prerecognition hypotheses for threatening words while not affecting hypotheses for control words. This differential stabilization of failure-word hypotheses seems to underlie both "vigilance" and "defense," suggesting that these response types are less distinctive than is usually assumed.

REFERENCES

1. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69-77.
2. CHODORKOFF, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, **49**, 508-512.
3. GLIXMAN, A. F. Recall of completed and incomplete activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281-295.
4. MCGINNIES, E. M., & ADORNETTO, J. Perceptual defense in normal and schizophrenic observers. *J. abnorm. soc. Psychol.*, 1952, **47**, 833-837.
5. MCGINNIES, E. M., & SHERMAN, H. Generalization of perceptual defense. *J. abnorm. soc. Psychol.*, 1952, **47**, 81-85.
6. NOWLIS, HELEN H., NOWLIS, V., RIESEN, A. H., & WENDT, G. R. Chemical influences on behavior. III. *ONR Tech. Rep.*, 1953, Proj. No. NR 144-060, 19 pp.
7. POSTMAN, L., & SOLOMON, R. L. Perceptual sensitivity to completed and incomplete tasks. *J. Pers.*, 1950, **18**, 347-357.
8. POSTMAN, L., BRONSON, WANDA C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215-224.
9. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.
10. WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
11. WHITTAKER, EDNA, M., GILCHRIST, J. C., & FISCHER, JEAN W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732-733.

Received March 19, 1956.

PRACTICE EFFECTS IN REACTION-TIME TASKS IN BRAIN-INJURED PATIENTS¹

ARTHUR L. BENTON AND HAROLD L. BLACKBURN

State University of Iowa

METHOD

REACTION-TIME studies of nonpsychotic patients with cerebral injury or disease have disclosed significantly slower times in both simple and choice reactions for such Ss when compared with appropriate control Ss. However, a considerable degree of overlap in the distributions of scores is typically found, the RTs of at least half of the brain-injured Ss being within normal limits (1). The question arises as to whether the presence of cerebral dysfunction in these Ss may not reflect itself in a diminished work-capacity, i.e., in a failure to maintain an initial level of performance in respect to speed and stability of response with continued activity. That such is the case has been asserted by Goldstein (2) on the basis of his observations on brain-injured soldiers during and after World War II. He reported that some patients showed initial simple and choice RTs that were well within normal limits. After the first ten trials, however, these patients were likely to show both a lengthening in RT and an increase in variability, whereas normal Ss did not show these changes within the limits investigated (30 trials). Thus, a relatively rapid decline in efficiency of performance, as reflected by slower and more variable RTs, appeared to be characteristic of many brain-injured patients whose initial RTs were unremarkable.

The purpose of the present study was to investigate this possibility of differential practice effects in brain-injured and control Ss somewhat more systematically than has been done heretofore. Matched groups of patients were run on both simple and choice RT tasks and the findings analyzed to determine what changes with respect to mean score and intra-individual variability occurred in the course of a series of 30 trials on each task.

¹ This investigation was supported by a research grant (B-616) from the National Institute of Neurological Diseases and Blindness, of the National Institutes of Health, Public Health Service. The writers are indebted to Dr. Robert J. Joynt for valuable suggestions and criticisms.

Subjects. The brain-injured group consisted of 38 patients (28 men, 10 women) on the neurological and neurosurgical services of the University Hospitals, Iowa City, who carried a final and unequivocal diagnosis of supratentorial cerebral disease. A variety of diagnoses was represented in the group, vascular disease and neoplasm being the most frequent (55 per cent). The mean age of the group was 36.9 years (*SD*, 12.0). Mean educational level was 10.0 years (*SD*, 2.6).

The control group consisted of 64 patients (52 men, 12 women) on the neurological, neurosurgical, and medical services who showed no evidence or history of cerebral disease, head trauma, or epilepsy. A variety of diagnoses was represented, spinal cord disease being the most frequent (42 per cent). Nine cases were classified as psychoneurotic with somatic complaints. The mean age of the group was 39.2 years (*SD*, 11.5). Mean educational level was 9.8 years (*SD*, 2.3).

No patient who was acutely ill, whose behavior raised the question of psychosis, who had a history of hospitalization for a psychiatric disorder, or whose history suggested mental deficiency dating back to childhood was included in the investigation.

As estimated by *t* and *F* tests, there were no significant differences between the two groups with respect to either the means or variances in age or educational level.

Procedure. The reaction-time apparatus provided for the presentation of a light stimulus which was preceded by a buzz of two seconds duration, the latter stimulus serving as a "ready" signal. Two lights, which were mounted seven inches apart on the right and left sides of an upright panel which faced the patient, were connected respectively to microswitches attached to the right and left sides of the base of the apparatus.

In the simple reaction-time task, the patient, having seated himself comfortably at the table, was instructed to rest the index finger of his right hand lightly on the right microswitch and to press the switch as soon as the right light appeared. The function of the buzz as a "ready" signal was explained to him. Five practice trials were given to acquaint the patient with the task. Thirty test trials then followed.

In the choice reaction-time task, the patient was instructed to rest the index fingers of his right and left hands on the right and left microswitches and to press the corresponding switch when the right or left light appeared. As in the simple reaction-time task, the light stimulus was preceded by a warning signal (buzz) of two seconds' duration. The patient was warned not to press both keys at the same time. Lights on the experimenter's side of the apparatus indicated the key which was pressed so that double pressing could be noted by the experimenter. Five practice trials were followed by thirty experimental trials, consisting of

fifteen stimulations on the right side and fifteen stimulations on the left side, presented in a predetermined random order.

The simple reaction-time task was given first and was followed, after a rest period of 2-3 minutes, by the choice reaction-time task. The reaction times were recorded automatically in .01 second units by an electrically driven clock which was activated by the onset of the light stimulus and stopped by the depression of the corresponding microswitch.

RESULTS

Simple RT. The means and standard deviations of the simple RTs in the two groups of Ss for successive blocks of 10 trials each are shown in Table 1A and the trends in mean score are graphically represented in Figure 1A. From inspection, it will be noted that while a difference in over-all RT level is obvious, there does not appear to be an impressive difference in the trends over blocks of trials, the control group showing no change and the brain-injured Ss showing a slight, questionably significant retardation in RT with continued practice. These impressions were confirmed by the results of an analysis of variance which showed a highly significant main effect for groups ($F = 25.7$ with df 1,100; $p < .001$) but no significant effects either for trials ($F = 0.59$, with df 2,200) or for the groups $\times$ trials interaction ($F = 1.11$ with df 2,200).² Thus it must be concluded that there is no evidence for significant practice effects, the two trends not being demonstrably different from each other or the combined trend significantly different from a horizontal line.

Three brain-injured Ss (8 per cent) showed an increase in RT of 20 per cent or more from the first to the third block of trials. The mean RTs of these Ss for the first block of trials were 197, 221, and 273 msec., i.e., two of them showed relatively fast initial RTs. Eight control Ss (13 per cent) also showed a retardation in RT of 20 per cent or more from the first to the third block of trials.

Choice RT. The means and standard deviations of the choice RTs in the two groups for successive blocks of 10 trials each are shown in Table 1B and the trends over trials graphically represented in Fig. 1B. It will be seen that while the control group showed a mean retardation of 3 msec. per trial between the first

² The designation, "nonsignificant," is meant to indicate that a result did not attain the .05 confidence level.

TABLE 1
REACTION TIME OF BRAIN-INJURED PATIENTS
AND CONTROLS FOR SUCCESSIVE BLOCKS
OF TRIALS
(Data in milliseconds)

Statistic	A. Simple RT					
	Brain-injured trials			Control trials		
	1-10	11-20	21-30	1-10	11-20	21-30
Mean	299	304	306	208	204	208
SD	149	141	139	41	45	47
Mean Intra-Individual SD	59	65	74	35	33	32

Statistic	B. Choice RT					
	Brain-injured trials			Control trials		
	1-10	11-20	21-30	1-10	11-20	21-30
Mean	371	382	389	303	304	306
SD	97	113	98	55	43	44
Mean Intra-Individual SD	71	73	70	55	49	49

and third blocks of trials, the brain-injured Ss showed a mean retardation of 18 msec. However, an analysis of variance indicated that this difference in trend, as estimated by the groups $\times$ trials interaction, was not significant ($F = 2.03$, df 2,200). As was the case with simple RT, a highly significant difference in the over-all performance levels of the groups was found ($F = 28.4$ with df 1,100; $p < .001$). The combined trend (trials effect) was not significantly different from a horizontal line ($F = 2.67$ with df 2,200). A separate analysis of variance of the brain-injured trend alone was done to determine if there might be a suggestion of a significant trials effect for this group. The obtained F was 2.29 (df 2, 74), indicating that the hypothesis that the trend represents a horizontal line could not be rejected.

Five brain-injured Ss (13 per cent) showed a retardation in RT of 20 per cent or more from the first to the third block of trials, the initial RTs of two of these Ss being relatively fast (i.e., under 330 msec.). None of these Ss was in the group who had shown noteworthy retardation in RT in the simple reaction task. One control S (2 per cent) also showed a retardation in RT of 20 per cent or more from the first to the third block of trials.

Levels analysis of performances of brain-injured Ss. Since clinical observation has em-

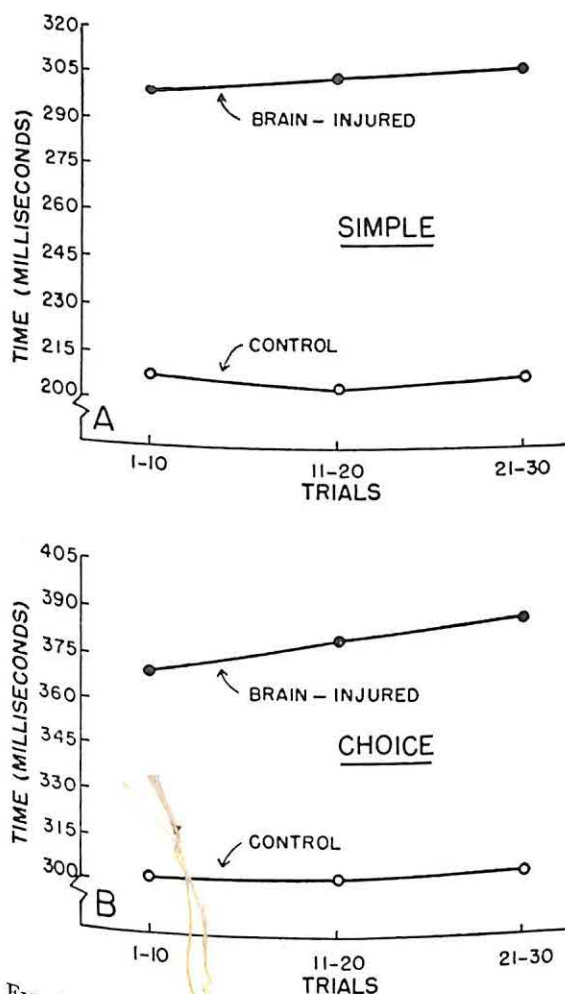


FIG. 1. TRENDS IN MEAN RT SCORES OVER BLOCKS OF TRIALS. A. SIMPLE RT. B. CHOICE RT

phasized susceptibility to fatigue in brain-injured patients whose initial RTs are within normal limits, it seemed indicated to analyze the data further to determine whether such Ss might show fatigue effects that were not apparent in the performances of brain-injured Ss with initially slow RTs. To make this analysis, the group of brain-injured Ss was reduced to an N of 36 by random subtraction of two cases and then divided into three subgroups of equal size ($N = 12$), defined by mean RT on Trials 1-10. This was done separately for the simple and choice tasks, with "fast," "medium," and "slow" subgroups in respect to mean RT on Trials 1-10 being formed by the counting-off method. The mean RTs over blocks of trials of these subgroups and the range of RTs for each subgroup on Trials 1-10 for the simple and choice tasks are presented in Table 2. Inspection of this table shows that,

apart from an expected slight "regression to the mean" in the extreme groups, there are no important differences in the trends of mean scores over blocks of trials, an impression that is supported by the results of analyses of variance of these trends which failed to show significant levels $\times$ trials interaction effects, the F for the simple and choice RT interaction being 1.04 and 1.54 respectively, each with df 4 and 66.

Changes in intra-individual variability. Since it has been asserted that brain-injured Ss tend to become more variable, as well as slower, with continued practice, this aspect of performance was investigated. Intra-individual variability was estimated by computing the mean SDs of the RTs for each block of 10 trials in each of the two tasks for each S . The mean SD for each block of trials for the members of each diagnostic group was then computed. These means are shown in Table 1. It will be noted that for both simple and choice RT the brain-injured Ss appear to be considerably more variable than the control Ss. Moreover, in the simple reaction task, the brain-injured Ss appear to show a progressive increase in intra-individual variability while the control Ss show no change in this regard. With respect to the choice reaction time no important differences in the trends of the two diagnostic groups are apparent. When these data were subjected to analyses of variance, highly significant group effects were found for both the simple ($F = 13.1$ with df 1,100; $p < .001$) and choice tasks

TABLE 2
REACTION TIME OF BRAIN-INJURED PATIENTS FOR
SUCCESSIVE BLOCKS OF TRIALS, ACCORDING
TO LEVEL OF INITIAL RT
(Data in milliseconds, range is for first
block of trials only)

Levels Subgroup	Simple RT		Choice RT	
	Range	Mean	Range	Mean
"Fast"				
Trials 1-10	154-221	195	245-325	297
Trials 11-20		208		292
Trials 21-30		217		310
"Medium"				
Trials 1-10	222-302	256	337-377	358
Trials 11-20		263		385
Trials 21-30		260		396
"Slow"				
Trials 1-10	307-876	466	377-686	477
Trials 11-20		459		480
Trials 21-30		460		475

($F = 17.4$ with df 1,100; $p < .001$). The trials effects (combined trends) were nonsignificant for both the simple ($F = 0.72$ with df 2,200) and the choice task ($F = 0.93$ with df 2,200). With respect to simple reaction time, the groups $\times$ trials interaction was significant ($F = 3.77$ with df 2,200; $p < .05$), indicating a difference in the trends of the two groups, the brain-injured *Ss* tending to become more variable than the controls with increasing number of trials. However, no difference in trends was found for the choice reaction task ($F = 0.81$ with df 2,200).

Nine brain-injured *Ss* (24 per cent) and nine control *Ss* (14 per cent) showed an increase of 100 per cent or more in intra-individual variability from the first to the third block of trials in the simple RT task. In the choice RT task, two brain-injured *Ss* (5 per cent) and one control *S* (2 per cent) showed an increase of 100 per cent or more in this characteristic of responsiveness.

DISCUSSION

The control *Ss* show a mean simple visual RT (207 msec.) over 30 trials which seems slightly slower than that which might be expected of healthy *Ss* but is still clearly within normal limits (3). No significant practice effects, either positive or negative, were shown over the three blocks of trials. While the brain-injured *Ss* were characterized by a mean simple visual RT (303 msec.) which is significantly slower than that of the controls, no difference with respect to trend of performance over blocks of trials was shown. Their performance, like that of the control *Ss*, did not appear to reflect significant practice effects. Moreover, this was as true of those brain-injured *Ss* who show fairly fast initial RTs (i.e., over the first block of 10 trials) as it is of those who show a retardation in RT from the very beginning. Thus, insofar as simple visual RT is concerned, no support is found for the expectation engendered by some clinical reports that brain-injured *Ss* may show an undue susceptibility to fatigue.

The findings for choice RT parallel those observed with respect to the simpler task. The brain-injured *Ss* were significantly slower than the controls but did not show a statistically significant differential practice effect. This held for those *Ss* whose initial RTs were relatively

fast as well as for the total group. Here, as in the simple RT situation, no substantial evidence for a greater susceptibility to fatigue on the part of brain-injured *Ss* was demonstrable.

The belief that brain-injured patients may become progressively more irregular in RT responsiveness with continued practice is partially supported by the findings. In the simple RT task the brain injured *Ss*, when compared with the controls, showed a significant increase in intra-individual variability over blocks of trials. However, this increase did not occur on the choice reaction task, in which no significant difference in trends was demonstrable.

While the analyzed results of the study are generally negative, it is still true that most of the findings were at least in the direction predicted by clinical report. In the simple RT task, the brain-injured patients showed a mean retardation of 7 msec. from the first to the third block of trials while the control patients showed no retardation. In the choice RT task the brain-injured patients showed a mean retardation of 18 msec. as compared with a mean retardation of 3 msec. in the control group. Moreover, in the simple RT task, a significant increase in variability with continued practice on the part of the brain-injured patients could be demonstrated. These results raise the possibility that modification of the task in the direction of making it more complex, lengthening the foreperiod beyond optimal limits, or increasing the number of trials might yield evidence of differential practice effects. This possibility can be decided only by further investigation.

However, while the study may provide only a limited test of the general hypothesis that brain-injured patients tend to show an undue susceptibility to fatigue in RT tasks, it does constitute a quite adequate test of the validity of the contention, derived from clinical impression, that such patients will show fatigue effects after the first few trials in simple and choice RT tasks. This does not prove to be the case. The findings do indicate that in each diagnostic group there are a few *Ss* who show a noteworthy retardation in RT or a striking increase in intra-individual variability from the first to the third block of trials in either the simple or choice task. It seems likely that clinical observers have been impressed by these changes in responsiveness on the part of some

brain-injured patients but have failed to consider the more numerous negative cases in the brain-injured group or the occurrence of the same type of change in the responsiveness of control patients.

SUMMARY

A study of performance trends in simple and choice reaction time tasks in nonpsychotic, brain-injured patients and a matched group of control patients indicated that, within the limits investigated (30 trials in each task), there were no significant practice effects with respect to mean reaction time for either group in either task. The performances of those brain-injured patients whose initial reaction times were within normal limits also failed to reflect significant practice effects. As compared with the control patients, the brain-injured patients showed a significant increase in intra-individual variability in simple reaction time with continued practice. However, this difference

in trends was not found in the choice reaction task.

The findings, while generally nonsignificant, raise the question as to whether a more complex task or a larger number of trials might not yield differential practice effects not disclosed in performance on these simple tasks over a limited number of trials. The results are interpreted as providing no support for the contention of some clinical observers that brain-injured patients are likely to show fatigue effects after the first few trials in simple or choice reaction-time tasks.

REFERENCES

1. BLACKBURN, H. L., & BENTON, A. L. Simple and choice reaction time in cerebral disease. *Conf. Neurol.*, 1955, **15**, 327-336.
2. GOLDSTEIN, K. *After-effects of brain injuries in war*. New York: Grune & Stratton, 1942.
3. TEICHNER, W. H. Recent studies of simple reaction time. *Psychol. Bull.*, 1954, **51**, 128-149.

Received March 21, 1956

A TEST OF THE TAYLOR-SPENCE THEORY OF ANXIETY^{1,2}

ELI SALTZ AND ARTHUR J. HOEHN

Air Force Personnel and Training Research Center, Chanute Air Force Base, Illinois

ONE of the most stimulating and provocative developments in recent psychology has been the attempt by Taylor and other Spence students to relate a clinically oriented definition of anxiety to modern behavior theory. Because of the potential significance of this development, it is important that the Taylor-Spence theory be rigorously tested. Unfortunately, some of the early experiments cited in support of the theory were inadequately controlled. The present experiments were designed to evaluate the Taylor-Spence theory under conditions which take into account some of the factors ignored in the early studies.

Taylor (10) originally conceptualized manifest anxiety as a drive variable. Her theory followed the conventional Hullian conceptualization of response strength (R) as a multiplicative function of a learning factor (H) and a drive factor (D). Since anxious Ss were theorized as operating at a higher drive level than nonanxious Ss , the prediction was made that more anxious Ss would learn faster than would less anxious ones. This theory worked out well in simple conditioning experiments; anxiety was found to facilitate the acquisition of conditioned responses (10). The theory proved inadequate, however, for more complex verbal learning. Montague (7) appears to have been the first to use the Taylor scale in a verbal learning study. He reported little evidence to indicate that anxiety facilitates verbal learning. In some conditions, he found that anxiety actually interfered with such learning. As a consequence of the Montague study, Taylor and Spence (11) have recently elaborated the earlier theory.

The revised Taylor-Spence theory holds that in a competing response situation, increased drive should increase the number of responses whose strengths are suprathreshold. Consequently, there should be an increase in the

number of erroneous responses which, through the operation of chance fluctuations in response strengths, interfere with the correct response. Since anxious Ss are assumed, after Miller (6), Mowrer (8), and Taylor (10), to have higher D than nonanxious Ss , anxious Ss should perform more poorly than nonanxious Ss in competing response situations.

Many studies have been performed that seem to support the Taylor-Spence theory (1, 7, 9, 11). However, in all these studies an increase in response competition is accompanied and confounded by an increase in difficulty level. In fact, at times difficulty seems to be used as an *index* of competition, as though no other sources of difficulty exist. For example, in one paper Farber and Spence state, "A further implication of the present theory was that the advantage of the nonanxious over the anxious group would be positively related to the level of difficulty of the individual choice points [of a finger maze], on the assumption that relative difficulty (number of errors) at a choice point reflects the strength of competing incorrect response tendencies at that point" (1, p. 122).

Because of the confounding of difficulty and competition in these studies, any theory based on the hypothesis that anxious Ss perform more poorly than nonanxious Ss as a function of the *difficulty* of the task would be upheld by the same data which have been used to date as evidence for the competing response theory.

The present studies attempt to control both competition and difficulty so as to determine their respective effects. In the first experiment, competing and noncompeting verbal learning materials were selected which had empirically been determined to have equal difficulty levels for a group of nonanxious Ss . The prediction, on the basis of the Taylor-Spence theory, was that the anxious Ss should do more poorly on the competing material than on the noncompeting, since the increased drive of the anxious Ss should increase the strength of competing erroneous responses. The second experiment attempted to compare the performance of anxious Ss on easy but competing material with their performance on difficult but non-

¹ This research was carried out under the Air Force Personnel and Training Research Center, Chanute Air Force Base, Illinois.

² The writers wish to express their appreciation to Professor Lawrence I. O'Kelly for his review of this research and his helpful suggestions concerning it.

competing material. The prediction from the Taylor-Spence theory was that anxious Ss should learn faster (relative to nonanxious Ss) when competition is reduced, even though difficulty is increased.

EXPERIMENT I

Method

Materials. The use of nonsense syllables permits the independent manipulation of response competition and the difficulty level of material. Both intralist competition and familiarity of the syllables contribute to difficulty in learning a list; greater competition increases difficulty, greater familiarity decreases difficulty. Consequently, it was possible to construct a list of familiar syllables with a great deal of intralist competitiveness which proved to be as difficult for a nonanxious group as a less familiar list with little intralist competitiveness.

The competing list was constructed from Glaze's (2) table of association values. An attempt was made to maximize competition in the following ways: (a) The first letter of each word was from the first half of the alphabet. (b) The same letter was used as middle letter for each word. (c) Three letters, each repeated three times, were used as the first letters of the list.

Lists at various levels of Glaze association value were tried before it was found that, for 10 nonanxious Ss who were counterbalanced as to the order of presentation of the two lists, a noncompeting list of 13 per cent association value was as difficult as the competing response list described above. The mean number of trials to a criterion of one perfect recitation was identical through the first decimal place for the two lists. The noncompeting list was constructed so that no two words in the list had the same first letter, and no two words had the same last letter. Letters were from the entire range of the alphabet.

Procedure. Groups of airmen were tested on the Taylor scale, and the upper and lower 20 per cent were selected for use in the experiment. Scores of 22 or above defined the anxious group; scores of 8 or below defined the nonanxious group. These cutting points are almost identical with those used in many of the Taylor-scale studies done at Iowa.

The Ss were randomly assigned to learn the two lists of nonsense syllables in the following manner. (a) The competing and noncompeting lists were each learned by a different group of 45 anxious Ss. (b) As a check on the previous equating for difficulty of the competing and noncompeting lists, 18 nonanxious Ss learned the competing list and 18 nonanxious Ss learned the noncompeting list.

All learning was by the anticipation method on a Missouri-type memory drum to a criterion of one perfect recitation. The syllables were presented at a 2-second rate with 8 seconds between repetitions of a list.

Results

The prediction, based on the Taylor-Spence theory, was that anxious Ss should do more poorly on material high in response competition

TABLE 1
MEAN NUMBER OF TRIALS TO CRITERION FOR ANXIOUS AND NONANXIOUS SUBJECTS ON "EASY" MATERIAL

Learning Material		Anxious	Non-anxious
Noncompeting 13% association material	Mean	31.09	27.83
	SD	13.59	14.64
	N	45	18
Competing 90% association material	Mean	28.09	28.44
	SD	17.54	12.18
	N	45	18

than on material low in response competition if the competing and noncompeting materials were of equal difficulty for nonanxious Ss. The results do not sustain this prediction. Table 1 shows the mean learning rates. With 45 anxious Ss in each group, $t = .907$ for the difference between competing and noncompeting lists; thus the null hypothesis can not be rejected. In fact, the difference is not even in the direction predicted.

The original equating of the competing and noncompeting lists for nonanxious Ss was cross validated with two new groups of nonanxious Ss. The results for these groups, summarized in Table 1, indicate that the initial equating of lists was satisfactory. The nonanxious groups had a difference between means of .61 trials to criterion; $t = .136$ with 34 degrees of freedom.

EXPERIMENT II

Method

Material. For this experiment it was necessary to find a noncompeting list which was more difficult than the competing list used in Experiment I. Since the competing list in Experiment I was as difficult as a noncompeting list at the 13 per cent Glaze association level, the difficult noncompeting list in this experiment was constructed from syllables at the zero association level. Rules for constructing this list were identical with those for constructing the 13 per cent association level noncompeting lists. As was the case in Experiment I, this list consisted of a nine-item series.

Procedure. Eighteen anxious and eighteen nonanxious Ss were run on the difficult, noncompeting material. These Ss were drawn at random from the same airman population used in Experiment I. Experiments I and II were run substantially simultaneously. The procedure for presenting the list was identical with that described in Experiment I.

Results

The scores obtained by the 45 anxious and 18 nonanxious Ss on the competing response list of Experiment I were used as a basis for evaluating the results of Experiment II. The

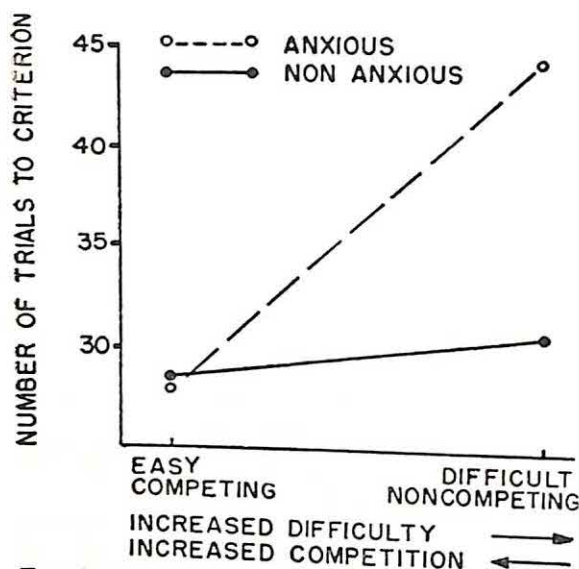


FIG. 1. THE PERFORMANCE OF ANXIOUS AND NON-ANXIOUS Ss WHEN LEVEL OF DIFFICULTY AND DEGREE OF COMPETITION ARE INVERSELY RELATED

prediction from the Taylor-Spence theory was that the anxious Ss should do relatively better (compared with the nonanxious Ss) on difficult, noncompeting material than on easy, competing material. The results are contrary to this prediction. As can be seen in Fig. 1, the deterioration of performance between the easy but competing and the difficult, noncompeting material was much greater for the anxious than for the nonanxious Ss. The difference between these mean differences is 14.50 trials to criterion, significant at beyond the .05 level with $t = 2.05$ and 95 degrees of freedom.

DISCUSSION AND CONCLUSIONS

Neither Experiment I nor II sustains predictions made on the basis of the Taylor-Spence theory. These experiments suggest that the results of previous studies which are often cited as support for the Taylor-Spence theory may be artifacts due to lack of appropriate control over the difficulty levels of competing and noncompeting materials. In Experiment I, the performance of anxious Ss was not worse on competing than on noncompeting materials, these materials having been equated on difficulty for nonanxious Ss. While the logic of statistics never permits proof of the null hypothesis, the indication that the anxious Ss perform no worse on the competing than on the noncompeting material is very convincing. The

N's are large (45 anxious Ss per group), and the difference is not even in the direction predicted by the Taylor-Spence theory.

In Experiment II, the results are also opposite in direction from those predicted on the basis of the Taylor-Spence theory, and this difference is statistically significant. A problem arises in connection with Experiment II, however. The noncompeting list of this experiment was assumed to be more difficult than the competing list of Experiment I. This assumption was made on the ground that the competing list was as difficult as a noncompeting list of 13 per cent Glaze association level; consequently, a noncompeting list of zero per cent should be more difficult than either the competing list or the noncompeting list at 13 per cent. As can be seen in Fig. 1, both anxious and nonanxious Ss do more poorly on the zero per cent noncompeting list than on the competing list, but this difference is not significant for the nonanxious S, though it is significant beyond the .01 level for the anxious Ss. Even if one assumes, however, that the noncompeting zero per cent list and the competing list are of equal difficulty, the results are contrary to predictions made from the Taylor-Spence theory. From these notions, one would expect the anxious Ss to do relatively more poorly on the competing material than on the noncompeting.

These findings give no conclusive idea as to how difficulty affects the performance of anxious Ss. However, some indication of how difficulty acts is suggested by other recent studies. Lucas (4) reported that the performance of anxious Ss deteriorates relative to the performance of nonanxious Ss when both groups are told that they are doing poorly on a task. Difficult material may trigger actual anxiety in manifestly anxious Ss, and this might somehow produce the decrement in performance. (It should be pointed out that the Taylor scale measures susceptibility toward anxiety, not necessarily the actual presence of anxiety in Ss. This point may be important for the interpretation of results obtained with the scale.)

Some recent research relating anxiety and intelligence appears to support the general conception presented here. Mayzner, Sersen, and Tresselt (5), reviewing the literature on the relationship between anxiety and intelligence, found that reported correlations ranged

between $+ .19$ and $- .40$. They presented evidence which suggests that no correlation between anxiety and intelligence occurs when the intelligence test is taken under nonstress conditions. These writers hold that negative correlations between anxiety and intelligence, such as those reported by Grice (3), may be due to the detrimental effect of stress on anxiety in the test-taking situation.

The above reasoning would suggest that increased difficulty might not be associated with relative decrements in performance by anxious Ss if (a) Ss were reassured of their good performance, or (b) Ss were in a state of stress while performing on the easy material.

SUMMARY

Previous studies, intended as tests of the Taylor-Spence response-competition theory concerning the effect of anxiety on performance, have confounded amount of competition and the difficulty level of the material. The present studies tested the Taylor-Spence theory under two conditions. (a) Competing and noncompeting materials were equated in difficulty for a nonanxious group. The prediction from the Taylor-Spence theory that the anxious Ss would perform more poorly on the competing than on the noncompeting material was not sustained. (b) Anxious and nonanxious Ss were tested on easy competing and difficult noncompeting materials. The prediction from the Taylor-Spence theory was that anxious Ss would do more poorly (relative to nonanxious Ss) on the easy competing than on the difficult

noncompeting material. The results were opposite from those predicted and significant at beyond the .05 level.

The results were interpreted as not supporting the Taylor-Spence theory.

REFERENCES

1. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120-125.
2. GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255-267.
3. GRICE, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, **50**, 71-74.
4. LUCAS, J. D. The interactive effects of anxiety, failure, and intra-serial duplication. *Amer. J. Psychol.*, 1952, **65**, 59-66.
5. MAYZNER, M. S., JR., SERSEN, E., & TRESSELT, M. E. The Taylor manifest anxiety scale and intelligence. *J. consult. Psychol.*, 1955, **19**, 401-403.
6. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89-101.
7. MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91-96.
8. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553-566.
9. RAMOND, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120-124.
10. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
11. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **42**, 61-64.

Received April 4, 1956

BREADTH OF DEVIATE CONCEPTS USED BY SCHIZOPHRENICS¹

LOREN J. CHAPMAN²

University of Chicago

AND

JANET A. TAYLOR

Northwestern University

IN attempting to account for the inferiority of schizophrenics in performing conceptual tasks, one hypothesis that has been advanced, most clearly by Cameron (1, 2), is that schizophrenics are able to form concepts quite readily but are prevented from exhibiting adequate performance by their overresponsiveness to inappropriate stimuli, both internal and external. Two recent studies by one of the present writers have lent support to this hypothesis (3, 4). In these experiments, Ss were required to sort cards by common qualities among figures on different cards, there being sufficient communalities between figures that the cards could be sorted in more than one way. In the instructions, *E* designated the "correct" communalities for sorting, thus making the task one of following instructions rather than of discovering the correct concept. The other communalities provided distracting opportunities for sorting in incorrect ways. It was found that schizophrenics, much more than normal Ss, tended to ignore the instructions and to sort by these distracting communalities. Further, as the number of these distracters in the test material increased, schizophrenic performance became increasingly inferior to that of normal Ss even though the tasks remained essentially the same. This distractibility was found both when the basis of the distracter communality was identity of the figures on the various cards and when the figures possessed communality by virtue of a conceptual category.

These studies showed that in one type of conceptual sorting task, at least part of the schizophrenics' inferior performance arises from overresponsiveness to inappropriate communalities, i.e., communalities other than the instructed ones. The present study investigates

¹ The authors are indebted to Drs. Julian H. Pathman and George Faibish of the Downey VA Hospital for assistance in obtaining the schizophrenic Ss, and to Drs. Gabriel Ferrazzano and Richard Saavedra of the Chicago USPHS Hospital for assistance in obtaining normal Ss.

² This study was done while the senior author was a postdoctoral research fellow of the National Institute of Mental Health.

interfering overresponsiveness in another kind of conceptual task and seeks to isolate other situational conditions under which it might occur.

It is commonly observed that schizophrenics tend to be "overinclusive" on conceptual sorting tasks, i.e., they include in a given category many instances which are incorrect. In addition, it is often stated that these incorrect instances are frequently related to or superficially resemble objects which properly belong to the category in which they are placed (5, 6). In the present investigation, certain aspects of this latter observation are subjected to experimental test. Specifically, it is hypothesized that the schizophrenic tendency to overinclude in a sorting task is positively related to the degree to which incorrect items share common qualities with the correct items in a given conceptual category.

The method of testing this hypothesis was as follows. Normal and schizophrenic Ss were told the name of a common conceptual category (e.g., fruits) and were asked to select from a group of items names of objects belonging to that category. A guide card containing the name of an object in the designated category served as a reminder of the specified category and also of the place where the correct items were to be put. The items from which Ss were required to select included, in addition to the correct instances, incorrect instances which were equally divided between members of a category similar to the designated one (e.g. fruits vs. vegetables), and members of a category that was markedly different from the correct one (e.g., birds). It was predicted that schizophrenics would incorrectly include in the designated category more similar items than dissimilar ones. In contrast, normal Ss, if they make overinclusive errors at all, should exhibit each type of error equally often.

Should the schizophrenics exhibit such a preference for overincluding similar items, it might be said that their basis for sorting tended to be inappropriately broad, being made according to a wider concept than was specified by the instructions. In the terminology that

will be used here, schizophrenics, when asked to sort by a given conceptual "species," would be sorting according to a broader "genus," the genus including not only the species specified in the instructions but also a second similar but incorrect species. In the example above, the genus of edibles, which included both fruits and vegetables, would be inappropriately substituted for the designated species of fruits.

In addition to being overinclusive, schizophrenics are also observed to be overexclusive in their sorting, i.e., to exclude from a category instances that properly belong to it. In attempting to account for this type of error, it is possible that such overexclusion results from sorting by a more narrow category than is appropriate. In the present study an attempt was also made to test this suggestion, adding a further hypothesis concerning the conditions under which such narrowing might occur. A summary of the actual experimental procedure might best serve to introduce these hypotheses. Using the same task materials as in the previously discussed sortings, *E* told *S* the name of a conceptual category that was, in the terminology used above, a "genus" and asked *S* to select from a group of items names of objects belonging to that category. The correct items available for sorting included equal numbers of members of two constituent species categories. The incorrect items were, as before, members of a category that was dissimilar from the genus or its two species categories. A guide card with the name of an object from the genus category was used to indicate the place for the genus cards. Naturally, the object on the guide card could belong to only one of the two species groups. It was expected that if errors of exclusion in schizophrenics are due to excessive narrowing of the conceptual basis of sorting, such narrowing would be most likely to occur in the situation in which a broad category, reducible to smaller ones, was employed. Further, in line with the notion that schizophrenics are highly distracted by external stimuli present in the situation, the direction of narrowing might be toward sorting according to the species to which the guide card belonged. Thus, schizophrenics were expected to exclude incorrectly from the genus sorting fewer items in the species category to which the guide card belonged than items in the species category to which it did not. Again it was anticipated that

such errors as normals might make would reveal no differential tendency toward the two types of overexclusion.

In one of the previously discussed studies of distractibility by Chapman (3), the criteria for choosing *Ss* were such that only moderately impaired schizophrenics could be used. Nonetheless, when the group was divided into two groups according to severity of illness, significant differences were found, the milder cases exhibiting fewer errors. The difficulty level of the tasks in the present study was such that more severe cases could be accepted. In order to determine whether degree of illness could again be related to the amount of performance impairment, two groups of schizophrenic *Ss* were tested, one judged to be relatively severe in degree of impairment and the other, comparable to the total group previously studied, moderately impaired.

METHOD

Subjects. One normal and two schizophrenic groups, each consisting of 24 males, were tested. The normal *Ss* were drawn from the general medical and surgical patients at the USPHS Hospital, Chicago, Illinois. No individual was included in this group who appeared by hospital record or by interview to have a psychiatric or neurological disorder. The schizophrenic groups, or neurological disorder. The schizophrenic groups, patients in the Downey VA Hospital, had all received a hospital diagnosis of chronic schizophrenia, with their first hospitalization for psychosis occurring at least one year previously and their last at least 3 months previously. Other prerequisites for selection were absence of any indication of cerebral organic involvement or mental deficiency, no current drug therapy, no EST within the past 3 months, or more than 35 EST treatments altogether. The two schizophrenic groups were distinguished by their ward status, the "moderately incapacitated" group being drawn from patients with ground privileges or assignment to an outside work detail, and the "severely incapacitated" group being drawn from patients with neither ground privileges nor outside work assignments. In addition to the 24 *Ss* in each of the schizophrenic groups, 4 moderate and 18 severe schizophrenic *Ss* were also seen but were not included in the final groups because they were unwilling or unable to participate in the experimental procedure.

Mean age was 37 years for each of the three groups, while mean years of education were 10.4 for the moderate schizophrenics, 10.6 for the severe and 9.8 for the normals. Although the groups did not differ significantly with respect to these two variables, both schizophrenic groups were significantly lower on functioning IQ as determined by the CVS, presumably because of the disruptive effects of schizophrenic pathology. On this test, the mean IQ was 107 for the normals, 88 for the moderate schizophrenics, and 92 for the severe.

TABLE 1
THE CATEGORIES AND GUIDE CARD USED IN EACH TASK, AND THE NAMES USED IN THE INSTRUCTIONS FOR THE CONCEPTUAL SPECIES AND GENUS

Task	Categories	Guide Card	Name for Species Category	Name for Genus Category
A	kitchen utensils furniture birds	pan	kitchen utensils, i.e., utensils used for cooking or eating	things found in a house
B	alcoholic beverages nonalcoholic beverages insects	wine	alcoholic drinks, i.e., drinks that if you drink enough of they make you tight	things that a person can drink
C	trees flowers musical instruments	oak	names of trees	growing things
D	fruit vegetables tools	apple	fruit	things that a person can eat
E	domesticated animals wild animals recreational equipment	cow	domesticated animals, i.e., animals that are kept on the farm or in the home	animals
F	wearing apparel household linens sea creatures	skirt	wearing apparel	things made of cloth

Materials. Six sorting tasks were devised, each consisting of 30 words typed on separate 3" × 5" index cards. The words were equally divided among three conceptual categories, two of which were designed to be similar but not overlapping (e.g., alcoholic and nonalcoholic beverages) and the third markedly different from the other two (e.g., insects). Each of the two similar categories is here called a "species" while the two together are called a "genus." The third, different, set of words is considered as control items. Before the experiment proper, the materials were given to a number of normal Ss, similar in background to those used in the experiment, in order to insure that for such Ss the items used were known and could readily be assigned to the proper category. The actual categories used in each of the tasks are listed in Table 1.

A 13" × 10" × 4" box was used for sorting the cards, the box having two slots in the top through which the cards could be deposited. A wire rack for holding a guide card was placed on top of the box in back of one of the slots.

Procedure. Each S was required to sort three of the six tasks according to a conceptual species, i.e., to separate the designated species cards from the second species and the dissimilar control items. The remaining three tasks were sorted according to a conceptual genus, i.e., separation of the genus category, made up of the two species, from the control items. Each of the six tasks could be used for either type of sorting, with only the instructions distinguishing the two.

For both types of sorting, E handed S the 30 cards of that task, named the category of cards that were to be separated out from the others and placed a guide card containing a word illustrative of the specified category on the rack in back of the slot in which cards of that category were to be placed. Instructions given on each task may be illustrated by the following, taken from the species directions for Task D (See Table 1): This card up here says "apple." An apple is a fruit. Take these cards and each one that names

a fruit goes in this slot beneath the apple. Everything else, that is, everything that is not a fruit goes in this slot here.

Except for the name of the particular category and the guide card used, these same instructions were used for all tasks on both types of sorting. The specific designations and the guide cards employed for both species and genus sortings are listed in Table 1.

Each of the six sorting tasks was used for species sorting with half of the Ss in each group and for genus sorting with the remaining. The order of presentation of the two types of sortings was of the abbaab pattern for each S, half of the Ss beginning with the species sort. The position of the guide card on the box was also varied for each S, alternating from right to left on successive tasks. Finally, the particular order of the six tasks was systematically randomized among Ss.

RESULTS

Species sorting. The distributions of all types of errors were extremely skewed, especially for the schizophrenic groups. For this reason, both mean and median error scores were computed for each group and all statistical tests were of the nonparametric type.

As predicted, the schizophrenics incorrectly placed in the designated species category more of the similar species items than dissimilar control words, while the normals made few errors of either type. (More specifically, as indicated by the median error scores shown in Table 2, the severe schizophrenics made 6 out of a possible 30 incorrect placements of the similar species items, the moderate group 6, and the normal Ss a median of zero errors. When compared by means of a Mann-Whitney

test, highly significant differences were found between schizophrenic and normal Ss, the comparison between the normal and moderate groups yielding a z of 2.76 ($p < .01$) and the comparison of normals and the severe schizophrenics a z of 2.78 ($p < .01$). No significant difference was found, however, between the two schizophrenic groups ($p = .43$). In contrast to these normal-schizophrenic differences for the similar items, the median number of errors of placement of the dissimilar cards was zero for all three groups and no significant differences were found among the groups.

One additional comparison was made in order to obtain a more direct measure of the tendency of the schizophrenics to overinclude differentially on the basis of similarity. For each S, the number of incorrectly placed dissimilar items was subtracted from the number of incorrectly placed similar items. As may be seen in Table 2, the median value of this difference score is zero for the normal group but much higher for the schizophrenic groups. (Examination of the data reveals that the lower difference score for the severe schizophrenics, as compared to the moderate group, is largely due to the tendency of the severe Ss to make more errors of *both* types, thus lowering the magnitude of the difference between the two types of errors.) Comparisons of the groups on these difference scores by means of a Mann-Whitney test yielded a z value of 4.10 ($p < .001$) between normals and moderate schizophrenics and of 3.44 ($p < .001$) between normals and the severe group. Again, the two schizophrenic groups did not differ significantly ($p = .74$), both showing an equal tendency to include incorrectly similar items. Inspection of the errors on each of the six tasks separately revealed that this differential error tendency of

the schizophrenics occurred on each of the tasks.

The schizophrenic groups differed from the normal group, it will be remembered, with respect to IQ scores, even though they were matched on education. Therefore the observed tendency of the schizophrenics to overinclude items differentially on the basis of similarity might be attributed to a lowering of functioning intelligence. In order to evaluate this possibility, 19 normal and 24 schizophrenic Ss were selected from the total groups in such a way that the mean and *SD* of the IQ scores for the two groups did not differ significantly. The mean IQ of these selected normal Ss was 105, and of the schizophrenics, was 104. When these groups were compared, the median score for the difference between the dissimilar and similar types of error was 0 for the normal and 2 for the schizophrenic Ss, while the means for the two groups were .1 and 4.8 respectively. The difference between the two groups, tested by a Mann-Whitney test, was significant at the .001 level of confidence, thus indicating that the tendency to overinclude is not accounted for by differences in functioning intelligence.

As a matter of incidental interest, attention might be called to a different type of error than those mentioned above: the number of correct species cards *excluded* from the designated category. From Table 2 it may be seen that the discrepancy between the normal group and each of the schizophrenic groups on this score was not as striking as on the over-inclusion measures, but in both cases the difference was significant at the .01 level.

Genus sorting. In the genus sorting, it will be recalled, correct sorting consisted of placing with the guide card all the members of a conceptual genus which consisted of two conceptual species. It was expected that schizophrenics might exhibit a differential tendency to *exclude* items from the designated genus category, i.e., to omit fewer of the items falling into the same species category as the guide card than items belonging to the second correct species. Contrary to expectation, however, the frequency of errors of exclusion was not related to the similarity of the items to the guide card. Although, as seen in Table 3, both schizophrenic groups excluded a few less of the species category to which the guide card belonged, the difference in frequency of the two

TABLE 2
MEDIAN AND MEAN NUMBER OF ERRORS ON THE SPECIES SORTINGS

Group		Inclusion			Exclusion
		Similar	Dissimilar	Diff. (S-D)	
Normal	Mdn.	0	0	0	0
	M.	.54	.33	.21	.58
Schizophrenic; moderate	Mdn.	6.00	0	6.00	1.50
	M.	7.33	.29	7.04	3.04
Schizophrenic; severe	Mdn.	6.00	0	3.50	1.00
	M.	8.20	1.33	6.87	2.67

TABLE 3

MEDIAN AND MEAN NUMBER OF ERRORS OF INCLUSION AND EXCLUSION OF GUIDE-CARD AND NON-GUIDE-CARD SUBCATEGORIES FOR GENUS SORTINGS

Group		Exclusion			Inclusion
		Guide-Card Cat.	Non-Guide-Card Cat.	Diff. (NG-G)	
Normal	Mdn.	0	0	0	0
	M	.25	.25	.00	.21
Schizophrenic; moderate	Mdn.	0	0	0	0
	M	.83	1.12	.29	1.42
Schizophrenic; severe	Mdn.	.50	1.00	0	0
	M	1.96	2.50	.54	2.00

types of error, examined by a Wilcoxon test, did not approach significance for any of the groups.

Errors of any type were, in fact, relatively infrequent in all of the groups on this genus sorting, most of the medians being zero. Further, when the groups were compared on each of the two types of errors of exclusion and on the errors of incorrectly including items from the third, incorrect category, none of the groups differed significantly. These figures, however, seem to overestimate slightly the goodness of the schizophrenics' performance. When groups were compared on total errors of exclusion, the severe schizophrenic group was found, by the Mann-Whitney test, to exceed the normal group significantly ($p = .02$). No other group comparison on this measure was significant, however. This general tendency toward exclusion that appeared in at least the severely incapacitated schizophrenics was paralleled, it will be recalled, by a similar tendency in the species sorting in which both schizophrenic groups significantly exceeded the normal Ss in their errors of exclusion. Thus, differences were found between normal and schizophrenic groups on overexclusion measures although the basis for this behavior has not been isolated.

The slowness of the differences between groups on the genus sorting, as compared with the differences on the species sorting, suggested that the genus task was less difficult for the schizophrenics than the species task. In order to check this more directly, the total number of errors made in each group on the two types of tasks were compared by a Wilcoxon test. Both schizophrenic groups were found to make more errors on species sorting than on genus sorting. For the moderately incapacitated, the z value

was 4.04 ($p < .001$) and for the severe, the z was 2.83 ($p < .01$). For the normal Ss, however, there was not a significant difference ($p = .73$).

DISCUSSION

The analysis of errors of inclusion in the species sortings confirmed the hypothesis that schizophrenics would incorrectly include more items from the category similar to the correct one than from the dissimilar category. In attempting to describe the overinclusive behavior of these Ss, it might be said that the schizophrenics continued to perform conceptually but according to a broader category than was specified by the instructions. These results thus add further support to the contention that schizophrenic inferiority on conceptual tasks is due to overresponsiveness to inappropriate stimuli rather than loss of conceptual ability per se.

The hypothesis that on the genus sortings schizophrenics would show the opposite reaction—excessive narrowing of the conceptual basis for sorting—received no support. Schizophrenics made more errors of exclusion than did normals (although only the severe group differed significantly from normal Ss), but the attempt to isolate stimulus conditions affecting these errors was unsuccessful.

One of the interesting results of the genus sorting in general was the fact that schizophrenics made so few errors of any type, inclusive or exclusive. In a sense the comparative adequacy of the schizophrenics' performance further confirms the findings of the species sorting: when faced with separating categories of items that are dissimilar (as in the genus tasks), schizophrenics differ comparatively little from normals; when the correct items are relatively similar to some of the incorrect, schizophrenic performance suffers although that of normals does not. Related to this is the unexpected finding that schizophrenics showed a more pronounced tendency to exclude correct items from the designated category in the species than in the genus sortings. A possible reason for this is that in the presence of stimuli that act strongly as distracters, schizophrenics become more susceptible to distraction by all sorts of other stimuli, internal and external, which under other circumstances do not have much influence.

Finally, a comparison can be made of the

two schizophrenic groups. The severely incapacitated group tended to make more errors of all types on both kinds of tasks than the moderately incapacitated and included more Ss with high error scores, as can be seen especially from the mean error values shown in Tables 2 and 3. Although the difference between groups in no case reached an acceptable significance level, there was, at least, a consistent direction of difference, the severely incapacitated Ss tending toward a more inferior performance.

SUMMARY

The present study attempted to isolate stimulus conditions that result in certain types of errors being made by schizophrenics on a conceptual sorting task. Normal and schizophrenic groups were asked to sort words naming objects falling into three conceptual categories, two of which were similar to one another, and the third dissimilar. In the first part of the study Ss were asked to separate items belonging to one of the similar categories from the remaining items. It was expected that for the schizophrenics, but not the normal Ss, the items incorrectly included would more often be from the similar incorrect category than from the incorrect dissimilar one. The results confirmed this expectation, schizophrenics differing significantly from normal Ss on a score of number of incorrect similar items included minus number of incorrect dissimilar items. These errors of the schizophrenics therefore may be described as resulting from an inappropriate broadening of the conceptual basis of sorting.

An attempt was also made, in the second part of the study, to determine whether an opposite kind of error, excessive narrowing of the conceptual basis for sorting, might also take

place under certain circumstances. The groups were again given items falling into three categories but the designated category which they were asked to separate from the other items consisted of the two similar subcategories. It was expected that schizophrenics, unlike normal Ss, might tend to sort according to only one of these subcategories, thus excluding incorrectly more of one than the other. This prediction was not confirmed, schizophrenics showing little difference in performance from normal Ss in these sortings.

A comparison was also made of the performance of severely and moderately incapacitated schizophrenics. Although the severe group was in no instance significantly different from the moderately incapacitated, the direction of difference between the groups pointed quite consistently to the relative inferiority of the severe group.

REFERENCES

1. CAMERON, N. Reasoning, regression and communication in schizophrenia. *Psychol. Monogr.*, 1938, 50, No. 1 (Whole No. 221).
2. CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, 34, 265-270.
3. CHAPMAN, L. J. Distractibility in the conceptual performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1956, 53, 286-291.
4. CHAPMAN, L. J. The role of type of distracter in the conceptual performance of schizophrenics. *J. Pers.*, in press.
5. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2 (Whole No. 239).
6. RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.

Received April 13, 1956.

CRITIQUE AND NOTES

EFFECT OF PREINFORMATION UPON HUMAN CHARACTERISTICS ATTRIBUTED TO ANIMATED GEOMETRIC FIGURES¹

RONALD E. SHOR

Brandeis University

THE present study attempts to demonstrate that the impression that a person forms of another individual's personality is partially determined by the characteristics of the other people with whom he is seen to interact. Following the work of Asch (1) and others (2, 3, 5, 6, 7, 8, 9), who have demonstrated that preinformation about an individual's personality affects first impressions of him, this study attempts to show that such preinformation also affects impressions of the personalities of others who interact with him. Animated geometrical figures were used instead of actual persons.

In the Heider-Simmel movie (4) which provided the stimulus material, three geometric figures (a large triangle, *T*, a small triangle, *t*, and a small circle, *c*) move about and interact in such a way as to lead most subjects (*Ss*) to report the action as that of three "persons" acting out a story. Most *Ss* spontaneously perceive *T* as an aggressive, "nasty" male who picks a fight with *t* and who attempts to take advantage of *c*; *t* is seen as a spirited, heroic male who attempts to protect *c* from *T*; *c* is seen as a fearful, defenseless female who seeks the protection of *t*.

It was hypothesized that if *Ss* were given different preinformation about the personality of one of the figures (*T* was chosen) the way in which the other two figures are perceived would also be affected. Since *Ss* tend spontaneously to perceive *T* unfavorably it was hypothesized that giving them unfavorable preinformation about *T* would either have no additional effect or would tend to accentuate slightly the way in which they spontaneously perceive the three figures. That is, if there were any changes, *T* would be seen as "nastier," *t* as braver, and *c* as more defenseless. On the other hand, it was hypothesized that favorable preinformation about *T* would cause a considerable change in interpretation. The *Ss* would perceive *T* more favorably and perceive *t* and *c* less favorably.

METHOD

In a pilot study 34 members of an undergraduate psychology course at the University of Kansas were

¹ This report is based on a dissertation submitted in May, 1955, in partial fulfillment of the requirements for the M.A. degree at the University of Kansas, under Dr. Fritz Heider, thesis advisor.

shown the Heider-Simmel movie twice and asked to describe the personalities of the three figures. Before seeing the movie, however, half of the *Ss* were told that *T* was an aggressive person disliked by most people, and the other half were told that *T* was a fair minded person liked by most people. From the *Ss'* reports the investigator listed all of the trait-names used to describe one or more of the three figures. This list was then reduced to 40 trait-names on the basis of frequency of report and diversity of coverage.

In the present experiment, 300 students from an introductory psychology course were shown the Heider-Simmel movie twice and then asked to check for each of the figures those of the 40 trait-names which were felt to be relevant. The conditions were identical for all *Ss* with the exception that 100 *Ss* received the aggressive-disliked preinformation about *T's* personality (AD group), 100 others received the fair-minded and liked preinformation (FL group), and 100 more received no preinformation, for purposes of control (Cn group).

The instructions for the AD group were as follows: "We are going to show you a movie in which there are three geometric forms; a large triangle, a small triangle, and a circle, moving in various ways in and around a larger square. Please imagine that the two triangles and the circle are people acting out a story. We will show you the movie twice and then ask you to write out the answers to a number of very simple questions about the movie. In no way is this a test of your intelligence or personality, and there is no time limit. The only bit of information about the movie which we are giving you is that the large triangle is an aggressive person who is disliked by most people. Remember this fact and use it to help understand and interpret the story: *the large triangle is an aggressive person who is disliked by most people.*"

Instructions for the FL group were identical except that the words fair-minded and liked were substituted for the words aggressive and disliked. In the Cn group the last two paragraphs were deleted.

RESULTS

In Table 1, the trait-names are arranged according to the size and direction of difference between the AD and FL groups in the frequency of their choice as descriptions for *T*. Since the size of each group is 100, these frequencies are also percentages. To illustrate, 92 *Ss* in the AD group checked the trait-name "not likable" for *T* in comparison to only 24 *Ss* in the FL group. The difference of 68 between these two frequencies is the largest in which an AD frequency exceeded an FL frequency, and thus "not likable" is ranked in

TABLE 1
FREQUENCIES FOR EACH FIGURE OF TRAIT
ASCRPTION IN THE AD AND FL
GROUPS

Trait-names	Figure					
	<i>T</i>		<i>t</i>		<i>c</i>	
	AD	FL	AD	FL	AD	FL
Not likable	92	24 *	1	23 *	2	9
Troublemaker	85	29 *	13	52 *	2	12 *
Bully	91	45 *	2	13 *	0	0
Bad person	57	15 *	1	18 *	0	6
Showoff	68	29 *	13	30 *	2	5
Treacherous	43	11 *	0	12 *	0	6
Defiant	48	22 *	21	31	5	8
Boastful	43	21 *	4	16 *	0	2
Unsure of self	48	26 *	18	31 *	55	52
Aggressive	90	70 *	23	35	4	5
Conceited	36	16 *	3	12 *	0	6
Cowardly	25	8 *	12	19	29	28
A tease	26	14	28	33	13	19
Hot-tempered	97	86 *	2	19 *	0	6
Tricky	23	12	20	38 *	2	8
Fickle	14	8	2	9	22	27
Curious	15	14	32	17 *	40	31
Defenseless	2	1	21	24	89	70 *
Timid	0	0	17	17	82	76
Dependent	13	14	15	17	63	65
Male	77	79	67	74	5	11
Persistent	66	69	44	44	4	4
Gullible	16	21	4	8	18	28
Female	4	10	9	8	76	72
Fun-loving	1	8	32	25	33	23
Minds own busi- ness	2	9	30	7 *	48	27 *
Peaceable	0	8	45	17 *	67	38 *
Clever	8	16	41	38	17	10
Playful	4	13 *	45	28 *	31	22
Honest	1	10 *	35	7 *	25	13 *
Loyal	0	10 *	59	28 *	31	29
Good natured	0	13 *	44	17 *	39	21 *
Likable	0	14 *	63	23 *	62	40 *
Courageous	5	23 *	59	31 *	8	3
Brave	4	24 *	61	33 *	6	3
Fair-minded	0	21 *	35	11 *	27	14 *
Good person	1	24 *	66	30 *	65	33 *
Has spirit	40	68 *	78	72	27	24
Protective	5	33 *	66	36 *	8	7
Easygoing except when provoked	14	58 *	43	17 *	11	6

* Difference significant at $p < .05$.

TABLE 2
AVERAGE FREQUENCIES OF FAVORABLE AND
UNFAVORABLE TRAIT ASCRPTION FOR
EACH FIGURE

Experi- mental Group	Figure					
	<i>T</i>		<i>t</i>		<i>c</i>	
	Fav.	Un- fav.	Fav.	Un- fav.	Fav.	Un- fav.
AD	6.2	63.5	54.5	8.7	28.3	7.6
FL	25.9	30.9	27.8	23.9	17.9	11.1
Cn	8.1	55.5	49.8	12.2	28.2	7.9

first place. An asterisk indicates that this difference is significant.² The trait-name ranked last is the one in which the greatest excess of an FL over an AD frequency occurred.³

As can be seen from the table, where more Ss in the AD group than in the FL group see *T* as possessing a certain trait, a very strong tendency exists for fewer AD than FL Ss to see this trait in the other two figures. The inverse is also true. Moreover, differences between AD and FL group frequencies are greatest for *T*, fairly large for *t*, and small but observable for *c*. Where the differences tend to be large for *T* they also tend to be large for *t* and *c*, and where they are smaller for *T* they tend to be smaller for the others also. Thus there is a clear and consistent tendency for impressions of *t* and *c* to vary together and in an opposite fashion from impressions of *T*.

The trait-names were dichotomized by the investigator according to whether they indicated favorable or unfavorable impressions of personality. In Table 2, frequencies for the unfavorable and favorable trait-names in which significant differences between the AD and FL Groups were found with respect to *T* are combined to yield a favorable and unfavorable average frequency for each figure for the three experimental groups.

In the Cn group where the stimulus configurations of the movie operate alone without any preinformation, *T* was viewed in a generally unfavorable light; *t* and *c* in a generally favorable light. These findings conform to expectations. In the AD group provided with unfavorable preinformation about *T* these same relations were preserved, and, except for *c*, were intensified. In the FL group, however, where the preinformation about *T* is contrary to the way most Ss spontaneously perceive the film, *T* was still more frequently seen as unpleasant than pleasant but many more Ss perceived *T* as possessing favorable traits and fewer perceived him as possessing unfavorable traits than in the AD group. The other two actors were also still seen as more pleasant than unpleasant

² Significance was accepted at the 5 per cent level. In most cases, however, probabilities were considerably less than the 1 per cent level.

³ Control (Cn) group frequencies are not included in this table. In most instances Cn frequencies are in between AD and FL frequencies, although much closer to the AD frequencies. To save printing costs, a table giving all of the data and the results of tests of significance for all three experimental groups has been deposited with the American Documentation Institute. Order Document No. 5034 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

but their favorable scores dropped and their unfavorable scores rose considerably.

Thus, within the limitations imposed by the compelling perceptual features of the Heider-Simmel movie, preinformation not only affected impressions of the figure about which it was given but also changed impressions of the other two figures. As impressions of *T*'s personality became more or less favorable, impressions of *t*, and, to a lesser extent, impressions of *c*, varied inversely. It is noteworthy that impressions of *c*'s personality changed less than did impressions of *t*. The major source of "human" interaction in the movie is seen as occurring between *T* and *t* with *c* remaining relatively passive. Since *c* interacts less with *T* than does *t* it is understandable why the different preinformation about *T* affects impressions of *c*'s personality less than it affects *t*'s.

SUMMARY

It was hypothesized that the animated geometric figures in the Heider-Simmel movie interact in such a way that the impression formed by *Ss* about the personality of one figure about which preinformation is given affects the way the other two figures are seen. This hypothesis was investigated by comparing the frequency with which favorable or unfavorable traits were ascribed to these figures under three conditions: favorable preinformation about one figure (100 *Ss*), unfavorable preinformation about this figure (100 *Ss*), and no preinformation (100 *Ss*). If it is legitimate to generalize from animated geometric

figures to real people in everyday life situations, the confirmatory results obtained tend to indicate that the impressions formed of an individual may be a function of the characteristics ascribed to another person seen to interact with him.

REFERENCES

1. ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258-290.
2. BERNALLES, I. The influence of family stereotypes in the judgment of personality. Unpublished master's thesis, Smith Coll., 1947.
3. GOLLIN, E. G. Forming impressions of personality. *J. Pers.*, 1954, **23**, 65-76.
4. HEIDER, F., & SIMMEL, MARIANNE. An experimental study of apparent behavior. *Amer. J. Psychol.*, 1944, **57**, 243-259.
5. HERSHKOWITZ, A. A further investigation into the processes of forming impressions of personality. Unpublished master's thesis, New School of Social Research, 1954.
6. KELLEY, H. H. The warm-cold variable in first impressions of person. *J. Pers.*, 1949, **18**, 431-439.
7. LUCHINS, A. S. Forming impressions of personality: a critique. *J. abnorm. soc. Psychol.*, 1948, **43**, 318-325.
8. MENSCH, I. N., & WISNER, J. Asch on "forming impressions of personality": Further evidence. *J. Pers.*, 1947, **16**, 188-191.
9. SIGURDARDOTTIR, V. An approach to the study of stereotypes of family roles and their influence on people's judgments of personality and behavior. Unpublished master's thesis, Smith Coll., 1946.

Received October 7, 1955

THE EFFECT OF STIMULUS CLASS ON THE PROBLEM-SOLVING BEHAVIOR OF SCHIZOPHRENICS AND NORMALS

ROBERT H. DAVIS AND ROBERT W. HARRINGTON¹

The RAND Corporation

ONE characteristic of the hospitalized schizophrenic is that generally he has been unable to achieve socially acceptable solutions to the problems of dealing with other individuals. Whether this inability is limited to certain classes of problems, i.e., those involving human stimuli, or indicates a generalized inability to solve prob-

lems, is of importance for personality theory. If for example, the inability of the schizophrenic to engage in effective problem-solving behavior is limited to problems involving human stimuli and not other classes of stimuli, it would be difficult to account for this fact with a theory of personality that stressed the importance of physiological status and failed to consider adequately the effects of interpersonal relations.

The present study was undertaken to ascertain the relative ability of schizophrenic and normal subjects (*Ss*) to engage in problem-solving behavior when the stimuli involved are of two types: human and nonhuman.

¹The present study was conducted while the authors were at Allegheny College, Meadville, Pa., and the VA Hospital, Marion, Ind., respectively. The authors wish to express their appreciation to Mary Jane Curry and Marie C. Morane both of whom participated in the collection and organization of the data, and to the staff of the Warren State Hospital, Warren, Pa.

TABLE 1
MEAN DIFFERENCE BETWEEN NORMALS AND SCHIZOPHRENICS MATCHED ON ONE OF THE TWO CLASSES OF
STIMULI AND COMPARED ON THE OTHER

Matching Variable	Comparison Variable	Matching Mean	N (Pairs)	Schizophrenic Mean	Normal Mean	Mean Diff.	<i>t</i>	<i>p</i> Sig. Level	Direction of Difference
Abstract Human	Human Abstract	3.57	45	3.25	3.78	.53	2.82	.01	N > S*
		3.43	44	3.69	3.40	.29	1.40	n.s.	S > N

* N.: Normal. S: Schizophrenic

METHOD

Although not a study in information theory per se, the approach used here to the measurement of problem-solving behavior was suggested by the work of Miller (2, 3) and others (1, 4, 5). The task involved the presentation of 32 pictures from which *S* was required to choose, on the basis of information supplied by *E*, the correct one. One set of pictures was heavily weighted with human content; the other contained only pictures devoid of human content.

Subjects. Fifty patients from the Warren State Hospital, Warren, Pennsylvania, and 50 students from Meadville High School, Meadville, Pennsylvania, served as *Ss*. Each patient had been diagnosed Schizophrenic Reaction and was judged by the hospital staff to be in good contact. No history of psychiatric disorder was present in the student population. Each patient was equated with a student on the basis of either IQ or grade achievement.²

Procedure. In both the human and nonhuman problems the pictures were selected so that when five bits of information had been given to *S*, he could locate any given picture out of the 32 arranged before him. At the outset of the study, standardized instructions explaining the test were read to *Ss*, and they were encouraged to ask questions. All questions were answered fully and, if necessary, the directions were repeated. During the problem, *Ss* were not told the extent to which they were correct or incorrect. Both groups were encouraged to listen closely and to do their best.

In the problem with human content, half of the pictures contained a male and half contained a female. The 16 male pictures were equally subdivided in the following manner: tie, no tie; book-in-hand, no book-in-hand; glasses, no glasses; and hat, no hat. The 16 female pictures were equally subdivided in the following manner: scarf, no scarf; book, no book; glasses, no glasses; and coat, no coat. When *S* was presented with the 32 cards containing human stimuli, he was given five bits of information, one bit at a time. Following the presentation of each bit, *S* was asked to indicate which of the 32 cards it might be. Thus, if the first bit was the statement, "It's a male," and if *S* selected a card with a male figure, he was credited with one bit. He might then be told that "the person for whom we are looking is wearing a tie." If *S* correctly indicated a male with a tie, he was credited with a second correct

bit. Then he might be informed that "the person for whom we are looking is carrying a book." Proper identification of a male with a tie and a book earned an additional bit of credit. This procedure was continued through five bits, and each *S* was given a score from zero to five, depending upon the number of bits of information about the cards which he could effectively utilize. The initial bit of information was alternately male or female. All *Ss* were given two trials with cards containing human content and two trials with cards containing nonhuman content.

The cards with nonhuman content were designed to parallel the human content cards and the problem was presented in the same manner. These cards, drawn with a Leroy Lettering Set, were divided into the following five bits: (a) printed in blue or red; (b) letters or non-letters; (c) if letters, big or little, and if nonletters, then numbers or figures; (d) printed on the top or bottom of the card; and (e) encircled or boxed in. Before presenting these cards to *Ss* the difference between letters and nonletters was explained to them. Numbers and figures were pointed out and the general differences noted.

RESULTS AND DISCUSSION

For the purpose of comparing the two populations, each schizophrenic *S* was matched with a normal *S* on the basis of his performance on the nonhuman task. These *Ss* were then compared by *t* tests for matched groups for performance on the problem involving human stimuli. When the groups were compared for performance on the nonhuman problem, a similar technique was used except for matching on the basis of performance on the task involving human stimuli. An inspection of Table 1 indicates that when *Ss* were matched on the basis of their performance on the nonhuman problems, a significant difference ($p = .01$) was found between schizophrenic and normal *Ss* in their ability to deal with the human pictures. When comparisons were drawn between the same two groups, matched for performance on human pictures, no significant difference existed between them in their ability to solve problems with non-human content.

It is apparent that the observed difference between normals and schizophrenics could be a function of either the stimulus class or the level of task difficulty. It is conceivable, for example, that

² No attempt was made to equate for age since previous exploratory studies by the authors indicated Pearsonian *rs* of very low order between performance on the task and age.

the inferior performance of schizophrenics on the human task was a function of the increased difficulty of the human task, independent of the stimulus class. There are two considerations which indicate that the crucial variable was stimulus class rather than level of problem difficulty. On the one hand, comparison of the performance of the normal Ss only on the two classes of stimuli indicates no significant differences with respect to mean scores of 3.57 (nonhuman) and 3.78 (human), and 3.43 (human) and 3.40 (nonhuman). Matched *t*s, computed for both sets of means, were not significant (1.08 and .15, respectively). If the two tasks differed in level of difficulty, one would expect this difference to be apparent in the performance of normal Ss on them. Normal Ss might not show such a difference, however, if both the abstract and human tasks were so easy that asymptotic scores were obtained on both. An inspection of Table 1 indicates that normal Ss did not obtain scores so near the upper limit (5.0) as to allow the conclusion that the tasks were exceptionally easy for either group of Ss.³

These results appear to indicate that the problem-solving ability of schizophrenics is a function of the class of stimuli with which they must deal and not necessarily a deficiency in problem-solving ability per se. Two such classes were employed in the present study. One of these classes involved human stimuli and the other nonhuman stimuli. Had the Ss been compared using other classes of stimuli, e.g., sexual symbols, mother figures, aggressive scenes, different racial groups, etc., the results might well have been different. In other words, a technique similar to the one used in this study might provide a method for isolating classes of stimuli which consistently and differen-

tially disrupt or impair the problem-solving ability of psychologically distinct groups of people, e.g., normal and psychotic; prejudiced and non-prejudiced; extroverted and introverted, etc. Such information could be used to test various theories of personality and could also provide important data to the practicing clinician.

SUMMARY

The ability of normal and schizophrenic Ss to deal with information involving human and non-human stimuli was compared. When Ss were matched on the basis of ability to utilize information about nonhuman stimuli, it was found that the problem-solving behavior of schizophrenics was significantly disrupted by the presence of human stimuli in the test situation. When the same populations were matched for ability to solve problems involving human stimuli, no significant difference was found in their capacity to deal with information involving nonhuman stimuli. The importance of this unilateral disruption of problem-solving behavior for personality theory was pointed out, and it was suggested that the technique utilized in the present study could be used to explore the effect of numerous classes of stimuli on normal and schizophrenic behavior.

REFERENCES

1. MILLER, G. A., & FRICK, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311-324.
2. MILLER, G. A. What is information measurement? *Amer. Psychologist*, 1953, **8**, 3-11.
3. MILLER, G. A. *Language and communication*. New York: McGraw-Hill, 1951.
4. ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183-193.
5. SHANNON, C. E., & WEAVER, W. *The mathematical theory of communication*. Urbana: Univer. of Illinois Press, 1949.

Received February 6, 1956.

³ Test-retest reliability coefficients were also calculated for these tasks. With respect to the human stimuli, the coefficients were .72 for normals and .66 for schizophrenics.

RESPONSE SET IN THE F SCALE¹

LOREN J. CHAPMAN²

University of Chicago

AND DONALD T. CAMPBELL

Northwestern University

IN the widely used F scale measure of authoritarian personality trends (1) all of the items are worded in a single direction, such that the "authoritarian" tends to agree with all items, the "nonauthoritarian" to disagree. Cronbach (4, 5) has marshalled considerable evidence to indicate that total scores on tests constructed in this fashion contain some systematic variance adhering not to the content of the items, but rather to a "response set" of "acquiescence," independent of specific item content. Consistent individual differences have repeatedly been found in the tendency to agree with items in general, and if not controlled in test construction, this "acquiescence" bias can generate sizable reliabilities within tests and correlations between tests that are independent of the reactions of persons to the specific content of the items. While, as Berg (2) suggests, acquiescence may in itself yield a useful personality measure, it often leads to a misinterpretation of findings when consistent individual differences are attributed to item content rather than to this more general trait. The standard precaution against such response bias in attitude research is to write items worded in both directions, so that the ideal high and low scorers must agree with half of the items and disagree with the others. The present paper reports on two efforts to so revise the F scale by reversing the wording of some F scale items.

In most attitude scales there is a specifiable social object presented, along with reactions to it, and it is a simple matter to present both favorable and unfavorable items. With the F scale the problem is much more difficult, and the present writers can make no defense of the particular reversals used. Take for example the item, "The wild sex life of the old Greeks and Romans was tame compared to some of the goings-on in this country, even in places where people might least expect it."

¹ This study was supported in part by the United States Air Force under contract No. AF 18(600)-170 monitored by the Crew Research Laboratory, Air Force Personnel and Training Research Center, Randolph Air Force Base, Randolph Field, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

² This manuscript was written while the senior author was a post-doctoral research fellow of the National Institute of Mental Health.

How is this to be reversed? One could say, "The sex life that goes on today even in places where one least expects it, is nothing compared to the wild sex life of the Greeks and Romans." Thus reversed, the item would still seem to appeal to the prurient moralizing that the initial item sought to tap. Better perhaps to state it, "Neither was the sex life of the Greeks and Romans, nor is the sex life of today as wild and immoral as is commonly thought." In the end we decided to leave this item alone.

In the first study we employed ten items in their original form³ and ten reversed items. Because these reversals are an essential part of the evidence, they are presented in full as the first 10 items in Table 1, which also indicates the test form and item number of the original item in *The Authoritarian Personality* (1) from which each was modified. In some analyses these two sets of items are employed as one 20-item test, but more frequently they are treated as separate 10-item tests, with the original wording being designated as F+, the reversals as F-. All items have been scored in the authoritarian direction.

In the first study, these 20 items were included in a longer questionnaire and administered to 134 students (69 male, 65 female) from the introductory course in psychology at Northwestern University, fall, 1953. Five levels of response were offered: strongly agree, agree, undecided, disagree, strongly disagree, but for scoring purposes the responses were dichotomized as agree and disagree, the undecided responses on each item being included with the less frequently chosen of the two. The F+ scale had a reliability⁴ of .69, the F- scale a reliability of .41. The correlation between them was positive and in the proper direction, but was only .17, significant at but the .05 level. To explore this failure in item reversal at the item-by-item level, tetrachoric correlations were computed between the F- items and the F+ total score. These values are also shown in Table 1. Note that the highest value, -.30, is in the wrong direction, indicating that the high-F+ person tends to agree with the item. As checked by χ^2 this relationship is significant at the .03 level. Whereas in the

³ These were items 1, 13, 19, 26, 31, 37, 38, 42 of Form 40, and 55 and 67 of Form 78.

⁴ All reliabilities were computed by an internal consistency method of Gulliksen (6).

TABLE 1
THE F- ITEMS AND THEIR CORRELATIONS WITH SCORE ON THE F+ SCALE

F+ Version*		F- Version	First Testing	Second Testing
Form	Item			
40	9	When a person has a problem or worry, he should try to think it through, rather than try to forget by keeping busy with cheerful things.	-.30	
40	2	It takes more than will power to conquer most obstacles.	-.11	
60	44	It interferes with learning and effective work when our teachers or bosses outline in detail what is to be done and exactly how to go about it.	-.10	
40	27	It would be better if people admitted to themselves that they do not always feel a great love, gratitude, and respect for their parents.	.12	.07
40	6	In spite of what some people say, human nature does not make war and conflict inevitable.	.18	-.06
40	21	It would be a good thing if more young people kept their youthful, rebellious ideas as they grew older.	.18	-.12
40	23	Strong leaders in whom people put complete faith are not what our country needs today.	.09	.12
40	34	Geeting rid of all of the immoral, feebleminded and crooked people would not actually solve many of the country's problems.	.28	-.06
40	41	The artist, writer and philosopher are probably more important to society than the business man and manufacturer.	.28	.10
40	43	Familiarity breeds respect more often than it breeds contempt.	.15	-.04
40	2	With most men it is luck and help from others rather than will power and drive that determines what eminence they attain in life.		-.16
40	8	It is unnecessary and undesirable for a person to have complete faith in some supernatural power whose decisions he obeys without question.		.06
40	12	Even though a person has bad manners, habits, and breeding, he should still be able to make himself acceptable to other people.		.02
40	16	It is completely absurd to believe that some people may be born with an urge to jump from high places.		-.03
40	18	Even though nowadays many different kinds of people move around and mix together a great deal, it is foolish to worry much about catching an infection or disease from them.		-.16
60	44	It interferes with learning and effective work when our teachers and bosses assume too active a leadership and tell us exactly what to do.		-.20
78	24	There have been other ages with more instability, conflict, and imminent danger than ours.		.07
78	38	It is a good thing to maintain the present emphasis in colleges on intellectual and theoretical topics, rather than change the emphasis to practical matters.		-.03

* For original items, see (1).

original wording the high-F person was anti-introceptive (When a person has a problem or worry, it is best for him not to think about it, but to keep busy with more cheerful things), by agreeing with the reversed wording (first item, Table 1) he appears introceptive. What is in common between the two wordings is that both are plausible declarative statements. Items 34 and 41 dealing with getting rid of the immoral and feebleminded, and the value of artists vs. businessmen, seem to have been more successfully reversed, and are now in a form in which the high-F person tends to disagree with them.

In the second effort at reversing F scale items,

16 F+⁵ and 15 F- items were employed. The F- items included the 7 of the first set which showed positive correlations, plus 8 new reversal attempts. These items are also shown in Table 1. These 31 items along with others were included in a questionnaire administered to 144 male students from introductory psychology courses at Northwestern University in the fall of 1954. The reliability of the F+ items was .53. The reliability of the F- items was .41. The correlation between the two scales was -.01. Again, tetrachoric correlations between

⁵ These were the 10 items used in the first study, plus Nos. 4, 9, 25, 33, 35, and 39 of Form 40.

the $F+$ total and the $F-$ items were computed. As can be seen from Table 1, the values are low and predominately negative.⁶ Although we have explored but a small sample of the ways in which these and other F scale items could be reversed, it seems fair to conclude that the achievement of a successful reversal would be a very difficult task.

It was possible in the second study to obtain a rather pure measure of the acquiescence response set through a special scoring of other questionnaire items administered at the same time. These items were intended to measure (a) superior orientation or a tendency to identify with one's superiors rather than with one's subordinates in situations involving cross pressures from these echelons; (b) alienation, or a tendency toward cynicism about the sincerity of other persons and conventional values; (c) identification with disciplining authority; and (d) preference for cooperation and doing things in groups. For each of these topics, an equal number of positive and negative items were constructed, to balance out the acquiescence response set. Rescoring the items for a general tendency to agree should thus produce a measure of the response set in which content factors were balanced out. Based on 74 items, this measure showed a reliability of .63. It correlated .43 with $F+$, -.06 with $F-$, and .32 with total F .⁷

DISCUSSION

If the F scale were free of response-set variance, one would expect $F+$ and $F-$ to correlate positively, to a degree approaching their reliabilities. On the other hand, if the systematic variance picked up by the F scale scores were only acquiescence response set, then $F+$ and $F-$ should show a high negative correlation. The zero or low positive correlation found indicates that neither extreme is true, and that the typical F scale score probably contains both response-set variance and variance adhering to the content of the items.

But without rejecting this conclusion we can note that it does not suffice for the data presented. On this basis taken simply, $F-$ should correlate as highly (though negatively) with the pure response-

set measure as does $F+$. It has about the same number of items, and all items are worded in one direction. But this outcome does not appear. It seems necessary to consider that the acquiescence response-set variance is correlated positively with the item-content variance of the F scale. The trait of agreeing with items seems to be symptomatic of a more general personality syndrome also symptomized by belief in the F scale content measured independently of the response set. Our best evidence of this is the highly significant correlation of .32 between the pure response-set measure and Total F in which response-set variance is presumably canceled out by the combination of items worded in both directions. On this basis, we would interpret the correlation of .43 between the response-set measure and $F+$ as due to the summation of the effects of response set in the $F+$ items plus content in the $F+$ items. The $F-$ items are so worded that these two sources of variance each potentially correlated with the pure set measure, subtract from each other and cancel each other out, resulting in the -.06 value. This interpretation provides in addition a rationale for the lower reliabilities of the $F-$ scales.

This interpretation by no means implies that the F scale will correlate only with test scores containing the response set. Ample evidence to the contrary is available, in which the F scale correlates significantly, if not highly, with measures of quite different construction (3). However, it does imply that the correlations with tests similarly constructed will be higher than they would be were the response-set bias controlled as by having items worded in both directions, by multiple choice format, or otherwise. This observation has important implications for some of the absolute levels of correlation reported in the initial study (1). Like the F scale, all of the Ethnocentrism items were worded in the same direction; thus the acquiescence response set must raise the correlation above what it would otherwise be. For the Political and Economic Conservatism scale, on the other hand, items worded in both the conservative and the liberal directions were employed. In this fashion, as well as in terms of the smaller number of items used, the dice were inadvertently loaded against finding political-economic attitude as important a correlate of ethnocentrism as the F scale measure of personality.⁸

⁶ The difference between the two studies in levels of correlation found may be due to the fact that the second study was based on males alone. In general, in our experience in this area, correlations and reliabilities run higher for women than for men.

⁷ In a further exploration, these items were split into 40 worded in the first person ("I feel," etc.) and 34 worded as declarative statements, as in the F scale. The acquiescence score for these had reliabilities of .46 and .47 respectively, correlated with each other .45, and correlated with $F+$.41 and .33. The difference in $F+$ correlations was not significant by the Pearson and Filon test (7), and was not in the expected direction.

⁸ Since this paper was prepared, two other F scale item-reversal studies have appeared: Leavitt, H. J., Hax, H., and Roche, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, 40, 215-221, and Bass, B. M. Authoritarianism or acquiescence, this Journal, 1955, 51, 616-623. Both

SUMMARY

An attempt was made to evaluate acquiescence response-set variance in the F scale measure of authoritarian personality trends by employing reversed wordings of items. Response set is found to be an important factor in the F scale which cannot easily be corrected by the practice usual in attitude-test construction of employing both positively and negatively worded items. Evidence seems to indicate in addition that the content component of the F scale is correlated with the acquiescence response set.

concur in finding evidence of response set variance in the ordinary form of the F scale. Bass confirms the lower F- reliability found here. Leavitt does not report reliabilities. Neither study provides the correlations of both F+ and F- with an external measure of response set which provides the major interpretive argument of the present paper.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J. & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BERG, I. A. Response bias and personality: the deviation hypothesis. *J. Psychol.*, 1955, 40, 61-72.
3. CHRISTIE, R. Authoritarianism re-examined. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of "The Authoritarian Personality"*. Glencoe, Ill.: Free Press, 1954. Pp. 123-196.
4. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, 6, 475-494.
5. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, 10, 3-31.
6. GULLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.
7. PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.

Received February 15, 1956.

A NOTE ON "ETHNOCENTRISM" AND ACQUIESCENT RESPONSE SETS

DOUGLAS N. JACKSON AND SAMUEL J. MESSICK¹

The Menninger Foundation

BOTH the California E scale and F scale (1) are composed entirely of positively stated items. The items are so worded that agreement with them indicates "ethnocentric" or "authoritarian" attitudes while disagreement presumably indicates the "equalitarian" end of assumed unidimensional scales. Recent investigations (2, 7, 8) have emphasized the importance of such a positive item form in using the F scale as a measure of attitude. These studies assessed the relationships between F-scale items in their original positively-stated form and F-scale items rewritten in "reversed" form so that disagreement rather than agreement could be considered an indication of authoritarian attitudes. The correlations between agreement to reversed and original items were generally found to be *positive*, i.e., people who agreed with positively-stated F-scale items also tended to agree with the reversed items even though the item content was opposite in direction. Thus, these studies indicate that when the accepted test practice of counterbalancing the direction of item content (9) is used in specifying an authori-

tarian scale, a paradoxical property of *negative reliability* is obtained.

This evidence obtained with the F scale may also be interpreted in terms of what Cronbach (4) has called an acquiescent response set, which has implications for interpreting the reported high correlations between the F scale and the E scale. There is a possibility that the observed covariation of responses to the two scales may be in part attributable to a consistent generalized tendency to agree or acquiesce to items in print, rather than being solely a function of consistent ideological beliefs.

In the present study, 93 junior and senior students² in high school, 40 males and 53 females ranging in age from 16 to 18, were given the 9-item E scale, Forms 40 and 45 (1), the reversed F scale, and the Gough intolerance scale (6). The groups were selected from social studies classes and contained only white, Christian students.

The reversed F scale was constructed by re-writing each item so that the content would appear to reflect a viewpoint opposed to the original, while retaining a similar style of expression. For

¹ This study was completed while the authors were USPHS Postdoctoral Clinical Research Fellows of the National Institute of Mental Health. Dr. Jackson is now at the Pennsylvania State University and Dr. Messick is at the Educational Testing Service and Princeton University.

² The writers are indebted to Mr. E. B. Weaver, Principal, Topeka High School, for kindly consenting to cooperate in this project.

³ The authors wish to thank Dr. Irwin Miller and James E. Simpson for help in reversing the items.

TABLE 1
THE REVERSED F SCALE

1. A love of freedom and complete independence are the most important virtues children should learn.
2. All the will power in the world will not help us when weaknesses and difficulties stand in our way.
3. The success of modern science proves that every important thing can ultimately be understood by the human mind.
4. Because human nature is improving, war and conflict will eventually be eliminated.
5. Every person should have complete faith in his own independent judgment, not in some supernatural power whose decisions he obeys without question.
6. When a person has a problem or worry, he should drop everything and concentrate upon it until the solution appears.
7. The deeper and more enduring qualities in getting along well with people are far more important than external things like manners, habits, and breeding.
8. What the youth needs most is complete freedom from discipline in order to flow naturally into the onward stream of cooperative progress for the benefit of all mankind.
9. It is known with complete certainty that the urge to jump from high places is learned, not inborn.
10. Nowadays, since democracy demands that people of widely different background and station mix together, a person should not be finicky about catching a disease from any of them.
11. An insult to our honor should always be overlooked, for "whosoever shall smite thee on thy right cheek, turn to him the other also."
12. The rebellious ideas that young people sometimes get must be encouraged and developed at all costs to guarantee mature citizenship in adulthood.
13. Only if prewar authorities are kept out of the German government, will true democracy be achieved in that country.
14. What this country needs most, more than willful leaders, are laws and political programs requiring all citizens to actively share responsibility.
15. Sex offenses, such as rape and attacks on children, never merit punishment; such offenders should always be treated with kindness and sympathy by qualified psychiatrists.
16. All attempts to divide people into the two distinct classes of the weak and the strong are doomed to failure.
17. Every truly mature person outgrows childish feelings of submissive respect and of excessive love and gratitude for his parents.
18. Such pseudosciences as astrology have never explained anything and never will.
19. The surest way to insure the disappearance of the true American way of life is to resort to the use of force.
20. In this scientific age there can be no justification for denying investigators the right to study so-called personal and private matters.
21. The complete end to wars and social problems will soon be realized.
22. There would be no immoral, crooked, or feeble-minded people, if we could get down to brass tacks and somehow get rid of our social problems.
23. The wild sex life of the old Greeks and Romans makes sexual goings-on in this country seem tame.
24. If people talked things over and didn't work so much, everybody would be better off.
25. It is foolish and ridiculous to have ideas that our lives could possibly be controlled by plots hatched in secret places.
26. Homosexuals are never criminals and must not be punished as such.
27. The business man and the manufacturer are not any more important to society than the artist and the professor.
28. No sane, normal, decent person can avoid having thoughts of hurting close friends or relatives.
29. We are bound to admire and respect a person if we get to know him well.
30. Nobody ever learned anything really important through suffering.

example, "Obedience and respect for authority are the most important virtues children should learn," was modified to read "A love of freedom and complete independence are the most important virtues children should learn," and "No weakness or difficulty can hold us back if we have enough will power," was changed to "All the will power in the world will not help us when weaknesses and difficulties stand in our way," etc. (see Table 1). There are probably alternative items that would better represent the "reverse" of the original statements. However, a defense of the adequacy of the reversals used in the present study would seem largely irrelevant if a certain typology like "authoritarianism" is to be defined in terms of belief in a particular ideological content and presumably disagreement with its reverse.

The Gough intolerance scale is a group of items empirically derived from the MMPI using the differentiation of extreme groups on the Levinson-Sanford Anti-Semitism scale as a criterion. The intolerance scale has been shown to correlate significantly with the California F scale, .46 and .43 (6). Incidentally, it should be noted with respect to response sets that 29 of the 32 items on Gough's scale are scored "true" to be in the "intolerant" direction. Gough (6) also reports a high negative correlation (-.73) between the intolerance scale and the MMPI K scale, a measure of test-taking attitude that contains 29 of 30 items scored false.

Cohn (3) has found a significant correlation between the F scale and a tendency to answer "true" to a diverse group of MMPI items and has

suggested that the correlation was due to a generalized tendency to agree with statements. Since the preponderance of items scored "true" on the intolerance scale suggests that it too may indicate some generalized tendency to acquiesce, we might expect such a response tendency to be reflected in a positive correlation between agreements to intolerance-scale and reversed F-scale items, even though there is a positive correlation between the intolerance scale and the original F scale. Of course, a positive correlation would be expected between the E scale and Gough's intolerance scale, because of their common empirical ancestry.

The results indicate that even though the Gough intolerance scale correlated positively with the E scale ($r = .24$, $p < .02$) in our sample, it also correlated *positively* with agreement to reversed F-scale items ($r = .23$). Thus, subjects who agreed with the intolerance-scale items tended to agree with reversed F-scale items. However, the fact that the intolerance scale appears to reflect an acquiescent response set may even contribute to its construct validity (5), but extreme caution should be exercised in defining the characteristics of respondents on the basis of item content. Also, the correlation between the E scale and the reversed F scale was .03, rather than being significantly negative as would be expected from item content. These results highlight the necessity of controlling item form in studying certain attitude contents, such as "ethnocentrism" or "authori-

tarianism." Also, previously reported relationships between beliefs and personality may take a different form when more adequate attitude measures are used.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.
3. COHN, T. S. The relation of the F scale to a response set to answer positively. *Amer. Psychologist*, 1953, **8**, 335.
4. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3-31.
5. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological testing. *Psychol. Bull.*, 1955, **52**, 281-302.
6. GOUGH, H. G. Studies of social intolerance: II. A personality scale for anti-Semitism; III. Relationship of the *Pr* scale to other variables. *J. soc. Psychol.*, 1951, **33**, 247-262.
7. JACKSON, D. N. Stability in resistance to field forces. Unpublished doctor's dissertation, Purdue University, 1955.
8. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215-221.
9. LOEVINGER, JANE. Some principles of personality measurement. *Educ. psychol. Measmt*, 1955, **15**, 3-17.

Received February 17, 1956.

A NOTE ON "DISCRIMINATION REACTION TIME AS A FUNCTION OF ANXIETY AND INTELLIGENCE"

JAMES T. FREEMAN AND HOWARD MAHER

Iowa State College

GRICE (3) has recently reported a functional relationship between intelligence and discrimination reaction time in accounting for results otherwise attributable to the operation of anxiety as a drive variable. It appears that the rationale for such an interpretation rests, to a great extent, upon his assertion that "... the discrimination reaction time test is known to be correlated with intellectual factors ..." (4). This statement does not appear to be completely in accord with the facts as reported in Grice's reference (4) and elsewhere (1).

On examination of Melton's data (4, pp. 1027-1030), one finds that discrimination reaction time (DRT) is only moderately related to an "intellectual" factor. Further, beyond the first trial in

DRT performance, intellectual components virtually disappear. On the other hand, such matters as perceptual speed and simple reaction time assume somewhat greater importance as performance progresses (1). Consequently any test involving perceptual factors may be expected to predict a perceptual criterion (DRT performance) reasonably well. Grice reported that the Clerical Aptitude Test, which he used as a "general intellectual measure," involves "perceptual factors" among others. Guilford (2) indicates that such "perceptual factors" correlate about as well as any others in predicting DRT performance.

In short, it would seem that Grice essentially demonstrated that a psychometric device with at least some weighting of perceptual factors does

reasonably well in predicting a type of performance that is largely perceptual in character! One may therefore question whether a functional relationship between *intelligence* and DRT has been reliably indicated and whether such a relationship is substantively likely. Finally, one may question the use of the Clerical Aptitude Test as the basis for equating groups for intelligence. Were not Air Force data on tests more universally recognized as general intellectual measures available?

It is believed that some of the points raised in the present note should have additional significance in view of the increasing use now being made of psychometric measures in the prediction and interpretation of experimental samples of behavior. There is implied in such an approach a joint responsibility among several research areas.

REFERENCES

1. FLEISHMAN, E. A., & HEMPEL, W. E., JR. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, **49**, 301-312.
2. GUILFORD, J. P., & LACEY, J. I. (Eds.) Printed classification tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 5.)
3. GRICE, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, **50**, 71-74.
4. MELTON, A. W. (Ed.) Apparatus tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4.)

Received February 27, 1956.

SOCIAL STATUS AND ABSTRACT BEHAVIOR

DONALD C. FINDLAY¹

*Human Research Unit No. 1
Fort Knox, Kentucky*

AND

CARSON MCGUIRE

University of Texas

IN recent years "intelligence tests" have been sharply criticized for an alleged cultural bias. It has been argued that this bias handicaps the bright child who has grown up in a lower socioeconomic environment. The present-day intelligence test, according to the argument, places a premium on knowledge readily accessible to the child in a middle-class home but seldom accessible to the child in a lower-class home.

Thus, when a lower-class child and a middle-class child are found to have similar IQs, it is possible that the lower-class child is actually "genetically" brighter, and that his performance was handicapped by the cultural bias of the test (3, p. 4). Presumably, if the two children were tested with intellectual problems involving materials with which they were equally familiar (2, p. 66), the lower-class child would probably make a higher score.

The present study is an attempt to evaluate this assumption by comparing children of similar IQ but of dissimilar socioeconomic background on a series of sorting problems requiring abstract behavior. The problems appeared to involve a knowledge of only four relatively simple block characteristics: height, width, shape, and color.

The hypothesis tested was: If lower-status children and middle-status children of similar

IQ are given block-sorting problems involving concepts of equal familiarity to the two groups, the performance of the lower status children will be significantly superior to that of the middle-status children.

METHOD

Subjects. Ninety-six subjects, selected by age-grade, index of status characteristics (ISC), and intelligence quotient, were used. Thirty-two Ss with IQs ranging from 90 to 110 were selected nonsystematically from each of three grades (the third, sixth, and ninth) in the public school system of a central Texas community. Half of the Ss at each grade level were selected from family backgrounds which rated as lower status (51 or less) by ISC standards. The other half of the Ss at each grade level were selected from middle-status (higher than 51) family backgrounds.

Age-grade refers to children of the same age at a given school-grade level. The index of status characteristics is the social class index developed by Warner and his associates (5). A modified form of the ISC based on occupation, dwelling area, and source of income was used (4). IQ was determined by means of the California Test of Mental Maturity. Third-grade Ss were tested on the Primary Short-Form, sixth-grade Ss on the Elementary Short-Form, and ninth-grade Ss on the Intermediate Short-Form. Differences between mean IQs of the two status groups at each age-grade were not significant (t 's of .11, .18, .13).

Materials. Four block-sorting problems similar to those of Ach (1) were used. In each problem, the S's task was to sort 20 blocks into four five-block groups. The four problems were derived from a matrix of

¹ This study was adapted from a doctoral dissertation submitted to The University of Texas by the senior author.

blocks representing combinations of four characteristics: height, breadth, color, and shape. There were two values of height (short and tall) and breadth (wide and narrow), and three values of color (blue, red, yellow) and shape (round, square, hexagonal). Blocks for a given problem were selected from the matrix so that only one sorting solution (basis of classification) was possible with the available values of the four characteristics.

The four sorting solutions, the four bases of classification, were always combinations of two characteristics (such as shape and breadth), each of which was present in two values (such as round and square) in the particular problem. In order to sort correctly, the *S* had to look at the 20 blocks and "abstract" the two characteristics which would permit a fourfold (2×2) classification. For example, in the problem based on shape and breadth, the correct solution consisted of sorting the blocks as follows: five round-wide blocks, five round-narrow blocks, five square-wide blocks, and five square-narrow blocks.

The other three problems (in addition to the shape-breadth [SB] problem) had solutions based on shape-height [SH], color-breadth [CB], and color-height [CH], respectively.

Procedure. The experiment was conducted in two phases: (a) by administering three block problems to an *S*; and (b) by checking the *S*'s knowledge of the simple concepts on which those problems were based.

In the phase *a*, the problems were presented in four orders: CB-CH-SB, CH-SH-CB, SB-CB-SH, SH-SB-CH. In each of the three age-grades, each of the four orders of presentation was given to four *S*s from each of the two socioeconomic levels.

Each problem was presented by pointing out the various characteristics and values of the blocks to the *S*, and then instructing him to sort out four "kinds" of blocks from the 20 before him. He was told that if his sorting were correct there would be an equal number of blocks of each kind. The instructions were phrased as simply as possible and in a manner intended to establish a "game" type of atmosphere. Performance was scored by using a system of "aids" which controlled the difficulty level of the problem.

An aid consisted of showing the *S* two blocks which illustrated the solution of the problem. The first aid presented the problem. The *S* had to note the characteristics in which the two blocks were similar, and then sort the rest of the blocks into the four categories on that basis. If the problem was not solved after a specified time, a second aid representing a second sorting category of the problem was used. The procedure was repeated until all blocks had been used as aids and the problem was automatically solved. An *S*'s score was the number of blocks which he needed as aids. Thus, the larger the *S*'s score, the poorer his performance.

Phase *b*, in which knowledge of the concepts used in the block-sorting problems was checked, took place immediately after the *S*'s response to the third problem. The procedure was to spread the blocks before him, hold up a block, and ask him to pick out all blocks which were the same as this block in the two characteristics on which one of his problems was based; i.e., "Give me all the blocks that are like this one in shape and height." The *S* was then given credit for knowledge of the concept if he picked out all the correct blocks. The

TABLE 1
SUMMARIZED ANALYSIS OF VARIANCE OF PERFORMANCE
ON BLOCK-SORTING PROBLEMS

Sources of Variance	df	Mean Square	F*	p
Order of presentation	3	1.66	2.72	—
Age-grade	2	5.26	8.62	<.001
Social status	1	5.21	8.54	.005
OP $\times$ AG	6	.96	1.57	—
OP $\times$ SS	3	.54	.88	—
AG $\times$ SS	2	.28	.46	—
AG $\times$ SS $\times$ OP	6	.46	.75	—
Within	72	.61		

* Within-group variance used as error term.

same procedure was followed in checking the concepts used in the other two problems.

RESULTS AND CONCLUSION

In order to determine the significance of the performance differences on Phase *a*, the block-sorting phase, between middle and lower status *S*s, analysis-of-variance procedures were used. Comparisons were made by age-grades III, VI, and IX, and by the four orders of problem presentation (Table 1). The mean difference between middle-status *S*s (12.4) and lower-status *S*s (15.2) was significant at the .005 level. The decrements in mean score from 17.1 for age-grade III, 13.1 for age-grade VI and 11.2 for age-grade IX were found to be significant beyond the .001 level. Differences between orders of presentation were not significant, nor were any of the interactions.

In Phase *b* of the experiment the chi-square test was used to compare differences between low- and middle-status *S*s' grasp of the simple concepts used in the sorting problems. None of the differences was significant. Frequencies of responses by social status (age-grade combined) for each concept yielded the following chi squares with 1 df: SB, .02; SH, .07; CB, 1.17; CH, .01. (Sig. significance at the .05 level requires a value of 3.8.)

The results of the Phase *b* confirm the assumption that the lower- and middle-status *S*s had about the same grasp of the concepts used in the block-sorting problems. Nevertheless, the hypothesis that lower-status children would excel in middle-status children on the sorting problems was not substantiated. Instead, it was found that the middle-status children were significantly superior.

DISCUSSION

This study does not support the argument that intelligence-type testing with "equally familiar" materials will expose the middle-class bias of standard intelligence tests. The two status groups here were equated in IQ as well as in the exper-

mental testing materials, yet middle-status Ss were clearly superior to lower-status Ss. Generalization from these findings is limited by the method of sampling (selection) used in obtaining Ss, as well as by the fact that only one intelligence test was represented.

The study does seem, incidentally, to support the belief that the ability to engage in abstract behavior increases with age and education through childhood and early adolescence. It is possible, however, that this finding resulted from successively more abstract content in the primary, elementary, and intermediate forms of the intelligence test used at the third, sixth, and ninth age-grades respectively to equate the two status groups.

SUMMARY

The contention that "intelligence" tests are biased against lower-class children was investigated by testing lower- and middle-status children of similar IQs with block-sorting (abstract behavior) problems. These problems involved only four relatively simple concepts found to be equally familiar

to the two groups. It was hypothesized that lower-status subjects are actually brighter than their IQs indicate, and that their performance on the block-sorting problems would be superior to that of middle-status Ss of similar IQ. The hypothesis was not confirmed; middle-status Ss were significantly superior to lower-status Ss.

REFERENCES

1. ACH, N. *Über die Begabungsbildung*. Bamberg: Buchner, 1921.
2. DAVIS, A. *Social-class influences upon learning*. Cambridge: Harvard University Press, 1950.
3. EELLS, K., et al. *Intelligence and cultural differences*. Chicago: Univ. of Chicago Press, 1951.
4. McGUIRE, C., & WHITE, G. *The measurement of social status*, Research paper in human development no. 3 (revised). Austin: The Univ. of Texas, May, 1952.
5. WARNER, W. L., et al. *Social class in America, a manual of procedure for the measurement of social status*. Chicago: Science Research Associates, Inc., 1949.

Received March 5, 1956.

HOW "RIGID" IS THE "AUTHORITARIAN?"¹

DOUGLAS N. JACKSON, SAMUEL J. MESSICK, AND CHARLES M. SOLLEY²

The Menninger Foundation

EVER since Rokeach (19) reported a relationship between "ethnocentrism" and "rigidity," a number of attempts have been made to demonstrate connections between ideological beliefs and personality variables. However, such theoretical relationships between ideology and personality are contingent upon the adequacy of the operational measures involved. The present authors question the validity of some of these measures, such as the F scale, for which the underlying dimension is defined in only one direction, i.e., scales for which agreements to positively-stated items designate one end of a continuum and disagreements with the same items presumably designate the other. The present research is an attempt to appraise the effect of such a scale form on previously observed

relationships involving "authoritarianism" and "rigidity."

As a parallel to Rokeach's study, Brown (3) substituted the California F scale for the E scale and found it necessary to introduce achievement-oriented conditions before it was possible to obtain a significant relationship between Einstellung rigidity and authoritarianism. French (8), however, did not obtain these results on a military population. The instability of this relationship has led to further investigation of the properties of the measures.

Both the validity and the generality of the Einstellung test as a measure of rigidity have been questioned (8, 9, 15, 18). Levitt and Zelen (15) suggest that the consistent use of set solutions in Einstellung problems may reflect conformity and overgeneralization. Pitcher and Stacey (18) concluded that submission may be the higher order construct under which rigidity can be subsumed. It may be possible to go one step further and relate the tendency toward conformity and overgeneralization elicited by the Einstellung situation to tendencies toward conformity revealed by agreement with certain types of positively-stated items, independent of content. Such a generalized tendency to acquiesce is a source of variation to which

¹ Portions of this paper were read at the Southwestern Psychological Association meetings in Dallas, 1956.

² This research was completed while the authors were USPHS Postdoctoral Clinical Research Fellows of the National Institute of Mental Health. Dr. Jackson is now at the Pennsylvania State University and Dr. Messick is presently at the Educational Testing Service, Princeton. The authors wish to thank Dr. Milton Rokeach for some thoughtful comments concerning the manuscript.

Cronbach (6, 7) has repeatedly called attention under the label of "response set."

Cohn (4), noting that all the items in the California scales are unidirectional, found a significant correlation between the F scale and a tendency to answer "true" to a diverse group of MMPI items. Jackson and Miller,³ Jackson (12), Bass (2), and Leavitt, Hax, and Roche (14), independently modified the direction of F-scale items so that disagreement rather than agreement would reflect the presumed authoritarian ideology. These studies yielded generally positive correlations between agreements with the original and reversed items. Such correlations do not indicate much about authoritarian ideology, but they do tell us something about the items themselves and suggest the use of a concept like negative reliability.

If this generalized tendency to agree with items independent of their content were adequately controlled, many of the relationships previously obtained with the F scale might be different (13). The present study is an attempt to interpret the reported correlation between the Einstellung test and the F scale as an indication that both measures are reflecting tendencies toward acquiescence, conformity, and overgeneralization, rather than as an indication that "authoritarians" should be labeled "rigid." Reversed F-scale items are also expected to reflect tendencies toward conformity and overgeneralization, such that the reversed F scale should correlate positively with the Einstellung test.

METHOD

Subjects

The Ss were 77 students from introductory psychology and history classes in a midwestern municipal university.⁴ Their median age was 20, their mean age was 23.5 and the SD 7.6 years.

Materials

The instruments used were (a) the Einstellung arithmetic problems, (b) California F scale, and (c) a specially constructed reversed F scale.

The Einstellung arithmetic problems. These problems were identical with those used by Brown (3). They consisted of two examples and seven items, each of which involved withdrawing certain amounts of water from a large jar in order to obtain a desired quantity. The second example and the two "set" problems could be solved by the formula $B - A - 2C$, where B is the largest jar and C the smallest jar. Following these were five problems which could be solved either by the longer formula above or by the shorter formula $A - C$. Brown noted that this procedure elicited "long" solu-

tions more frequently than the one used by Rokeach (19).

The original and reversed F scales. In addition to the 30 item California F scale, Forms 40 and 45 (1), Ss were administered a reversed F scale,⁵ which was constructed by reversing the direction of the item content while retaining as much of the original style as possible. The reversed F-scale items are reported by Jackson and Messick (13).

Procedure

Two weeks prior to the administration of the Einstellung test and the reversed F scale, approximately half the Ss were given the original F scale. About two weeks after taking the Einstellung test and the reversed F scale, the original F scale was administered to most of the remaining half.

Since Brown observed that an ego-involving atmosphere was essential for obtaining correlations between "authoritarianism" and "rigidity," a modification of McClelland's procedure (16) was used to create achievement conditions. The experimenter, dressed in a dark suit, was introduced to the class as a psychologist who was going to administer some very important tests. Instructions were then read which suggested that the tests were measures of intelligence and leadership potential and that the Ss were being compared with students from a nearby state university. They were instructed to write their names on the test papers and then to proceed with the Einstellung test. After the forms were completed, they were given some information about the true nature of the test in order to relieve any anxiety that may have arisen as a result of the achievement-oriented conditions.

RESULTS

Of the 77 Ss to whom the Einstellung test was administered, ten made one or more errors in arithmetic and were therefore eliminated from the sample. Of the 67 Ss remaining, 13 who failed to use either the set or the critical solution consistently were also eliminated. This left 32 Ss who always used the long solution and 22 who always used the short solution.

Two single classification analyses of variance were computed between these two Einstellung groups, one comparing them with respect to original F-scale scores and the other with respect to reversed F-scale scores. Since directional hypotheses were made, one-tail tests were used. Table 1 indicates that both the F scale and the reversed F scale differentiate the Einstellung groups at the .05 point. Thus, Ss who agree with F-scale items tend to use the set solution on the Einstellung test. However, agreement with reversed F-scale items for the same Ss also tends to covary with the use

³ Unpublished study, Purdue University, 1953. Discussed in (13).

⁴ The authors are indebted to Dr. Bernard Spilka for making the F-scale data available and to Drs. Spilka and Homer Rutherford for providing class time.

⁵ We wish to thank Dr. Irwin Miller and James E. Simpson for help in constructing this scale. For the 42 Ss taking the original F scale, mean item score was 3.10; SD, .67. The 54 Ss taking the reversed scale received a mean item score of 4.05; SD, .46.

TABLE 1
DIFFERENCES BETWEEN EINSTELLUNG GROUPS
WITH RESPECT TO ORIGINAL AND
REVERSED F-SCALE SCORES

F-Scale Scores					Reversed F-Scale Scores				
Source	df	Mean square	F	p	Source	df	Mean square	F	p
"Long" vs. "Short" Solvers	1	1227	3.11	<.05	"Long" vs. "Short" Solvers	1	539	2.84	<.05
Within	40	394			Within	52	190		

of the long Einstellung solution. Thus, by changing the direction of the item content, one may obtain a "rigid" performance from people giving either "authoritarian" or "nonauthoritarian" responses.

The correlation between agreement with original and reversed F-scale items was $+ .35$, rather than being significantly negative as consistent responses to content would require. The corrected split-half reliability of the reversed F scale was $.77$. The counterbalanced order of presentation of the original and reversed F scales contributed no significant effects. Also, the Einstellung results for both the F scale and reversed F scale were in the predicted direction for each of the two orders of presentation.

DISCUSSION

The results emphasize the importance of controlling for response sets in constructing attitude scales. It seems reasonable to assume that agreement with a statement can be interpreted as reflecting belief in its content only if disagreement with the opposite content can also be obtained. The tendency to agree with an item, *any* item, especially when written in a cliché-like fashion may reflect important personality characteristics. These in turn may manifest themselves by increasing the correlational validity of the scale but not necessarily the validity as suggested by the content of the items (7, 11). When a respondent is confronted with an item which contains a familiar phrase to which he has given little thought, agreement may represent merely a passive and acquiescing response. A preponderance of responses indicating agreement to a large number of diverse items could thus result either from belief in all of the statements or from an acquiescent response set. Similarly, disagreement with an item may or may not reflect disbelief in its content; a person may be objecting merely to the style in which it is written.

The present study emphasizes the importance of acquiescence in relationships between the F scale and the Einstellung test. It is important to

note, however, that Rokeach's (19) original findings were with reference to the *E scale* and Einstellung. Thus the conclusion that Rokeach's results are primarily attributable to acquiescence or response set may not be warranted by the present evidence. It would be valuable to repeat Rokeach's study substituting a different measure of anti-minority attitude, such as a social distance scale, to determine if this also correlates with problem solving behavior.

We do not mean to imply that the search for personality and cognitive correlates of attitudinal clusters is fruitless or that relationships between rigidity and attitudes are inconceivable. The work of Stern, Stein, and Bloom (20) indicates that when each end of an authoritarian scale is defined by both positively- and negatively-stated items, important relationships still appear, and in such a case it is possible to separate belief from response set.

Thus, if one end of an attitude continuum is defined by agreement with positively-stated items, as is the case with the F scale, it does not follow that the opposite end of the continuum can be designated by merely tallying disagreements to the same items. It should also be noted that even if the direction of item content had been controlled in constructing the F scale, the mathematical model implicit in a Likert scoring procedure requires a homogeneous scale. As Green (10) has pointed out, "the intercorrelations among the items are assumed to be due to a single common factor to which the items are all mutually related." Since there is evidence for multidimensionality in the F scale (e.g., 5), a revision more basic than just controlling for acquiescent response sets is probably necessary before it can justifiably be used for measuring an attitudinal dimension (10, 17).

SUMMARY

This study is an attempt to appraise the role of a generalized tendency to agree with items independent of content in observed relationships between "authoritarianism" and Einstellung "rigidity." Subjects were given the California F scale, a reversed F scale, and a series of Einstellung water-jar problems. Subjects who used the long "rigid" solution in the Einstellung situation tended to agree with both F-scale and reversed F-scale items. These results suggested that both the Einstellung test and the F scale may be reflecting acquiescence and conformity. Also, the importance of defining both ends of a scale with positively-stated unidimensional items before drawing conclusions relating attitudes and personality was emphasized.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 611-623.
3. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469-476.
4. COHN, T. S. The relation of the F scale to a response set to answer positively. *Amer. Psychologist*, 1953, **8**, 335. (Abstract)
5. CORDER, R. F. A factorial approach to anti-democratic attitudes. *Studies in higher education*. Lafayette, Ind.: Purdue Division of Educational Reference, 1954.
6. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475-494.
7. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3-31.
8. FRENCH, ELIZABETH G. Interrelation among some measures of rigidity under stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 114-118.
9. GOODSTEIN, L. D. Intellectual rigidity and social attitudes. *J. abnorm. soc. Psychol.*, 1953, **48**, 345-353.
10. GREEN, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Vol. 1. Pp. 335-369.
11. GULLIKSEN, H. Intrinsic validity. *Amer. Psychologist*, 1950, **10**, 511-517.
12. JACKSON, D. N. Stability in resistance to field forces. Unpublished doctor's dissertation, Purdue Univ., 1955.
13. JACKSON, D. N., & MESSICK, S. J. A note on "ethnocentrism" and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132-134.
14. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215-221.
15. LEVITT, E. E., & ZELEN, S. L. The validity of the Einstellung test as a measure of rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 573-580.
16. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
17. MESSICK, S. J. Some recent theoretical developments in multidimensional scaling. *Educ. psychol. Measmt*, 1956, **16**, 82-100.
18. PITCHER, BARBARA, & STACEY, C. L. Is Einstellung rigidity a general trait? *J. abnorm. soc. Psychol.*, 1954, **49**, 3-6.
19. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259-278.
20. STERN, G., STEIN, M., & BLOOM, B. *Methods in personality assessment*. Glencoe, Ill.: Free Press, 1956.

Received March 9, 1956

ANALITY, AGGRESSION, AND ACQUIESCENCE

WILLIAM RABINOWITZ

Bank Street College of Education¹

FARBER (4) has recently presented evidence of a relationship between anality and political aggression as hypothesized in psychoanalytic theory. He found a significant correlation of .37 between two tests, the first a five-item measure of anality and the other a four-item measure of political aggression. An examination of the items in the two tests shows that the correlation may be a function of what Cronbach has termed "response set" (2, 3) rather than the result of a relationship between anality and aggression. On each of the anality items, the "Yes" option was keyed positively, and on each of the political aggression items, the "Agree" and "Strongly Agree" options were keyed positively. The keying of the items is such that scores on the two tests are influenced in the same way by any general tendency on the part of subjects to acquiesce, i.e., to say "Yes," "Agree," or "Strongly Agree," when responding to test

items. There is ample evidence that response sets, such as acquiescence, are a reliable source of variance in test scores (2, 3), including tests of attitudes (1, 6), interests (5), and personality characteristics (5). In view of this evidence, there is reason to believe that Farber's results may stem from acquiescent tendencies on the part of subjects, rather than an inherent relationship between anality and aggression.

The most appropriate method of studying the anality-aggression relationship is to employ tests in which response sets cannot systematically influence scores. In the absence of such instruments, an alternative procedure is to add to each of Farber's tests an equal number of items which measure anality and political aggression, but which are keyed so that the "No" and "Disagree" options are positive. By examining the correlations among the original Farber tests and those newly constructed, it should be possible to determine whether the reported anality-aggression relation-

¹ Now at the Division of Teacher Education, Municipal Colleges of New York.

ship is an artifact of the way in which the test items were keyed.

Using the concept of anality as summarized by Farber, several new test items which seemed to embody this trait were constructed. Unlike Farber's items, however, the items were stated so that a "No" response would signify greater anality. The five items follow:

I usually lend my possessions to people who ask permission to use them.	Yes	No
If I'm doing something I enjoy, I often stay up late at night, sometimes so late that I don't get a full night's sleep.	Yes	No
I'd rather lead a life of excitement and adventure than one which is stable, predictable, and well-organized.	Yes	No
Some people think it's a good idea to keep a budget so that purchases can be carefully planned, but in my case I don't think I could stick to a budget for very long.	Yes	No
When I go to a restaurant I like to order foods which I have never tasted before.	Yes	No

Henceforth, Farber's anality test will be referred to as the A_1 scale, the author's anality test as the A_2 scale, and the combined 10 items as the A_T scale.

Following Farber's concept of political aggression, a test consisting of four items was built to measure this attribute. Here too the items were so stated that a "Disagree" response would signify greater aggression. The author's four items follow:

There is good reason to believe that the nations of the world can live in peace.	Agree	Disagree
The foreign policy of the United States should be based on the assumption that most other countries are our friends, not our enemies.	Agree	Disagree
Capital punishment should be abolished.	Agree	Disagree
It is unhealthy in a democracy to suppress minority viewpoints no matter how distasteful they may be.	Agree	Disagree

Henceforth, Farber's test of political aggression will be referred to as the P_1 scale, the author's test of political aggression as the P_2 scale, and the combined 8 items as the P_T scale.²

² In Farber's study each of the political aggression items had four response options: Agree Strongly, Agree Slightly, Disagree Slightly, and Disagree Strongly. The simpler procedure of using only an Agree and Disagree option for each item was used in this study, because this eliminated the influence of another response set, i.e., willingness to go to extremes in registering agreement or disagreement.

TABLE 1
CORRELATIONS OF A_1 , A_2 , P_1 , AND P_2
($N = 143$)

	A_1	A_2	P_1	P_2
A_1	+.30†	+.12	+.18*	-.04
A_2		+.32†	-.05	+.15*
P_1			+.53†	+.16*
P_2				+.25†

* .01 < p < .05.

† Reliability coefficients based on Kuder-Richardson Formula 20.

The A_T and P_T scales were administered to a group of 143 undergraduate college students (approximately 90 per cent female) who responded anonymously. Both scales were administered in a single session, the P_T scale immediately following the A_T scale. The responses given by the subjects were summarized in the following scores:

	Test	Way of Obtaining Score
Anality scores	A_1	Number of "Yes" responses
	A_2	Number of "No" responses
	A_T	Sum of above two scores
Political Aggression scores	P_1	Number of "Agree" responses
	P_2	Number of "Disagree" responses
	P_T	Sum of above two scores
Acquiescence scores	A_T	Number of "Yes" responses
	P_T	Number of "Agree" responses

On the assumption that anality and political aggression are more closely related than a response set to acquiescence, the correlation of A_T and P_T should be higher when scored for anality and political aggression than when scored for acquiescence. The actual correlations are as follows:

A_T vs. P_T (scored for anality and aggression)
... $r = +.09$ (not significant)

A_T vs. P_T (scored for acquiescence) ... $r = +.27$
(significant at the .01 level)

These correlations suggest that the reported anality-aggression relationship is a function of acquiescence. Further evidence to support this view appears in the intercorrelations of A_1 , A_2 , P_1 , and P_2 as presented in Table 1. Both the correlation of A_1 with P_1 , and A_2 with P_2 reach significance, though Farber reported a higher correlation between A_1 and P_1 . Since these correlations involve tests whose items are scored in the same direction, it is impossible to tell whether they are a function of acquiescence or an anality-aggression relationship. This ambiguity is not present in the correlations of A_1 with P_2 and A_2 with P_1 , since the items

of the tests involved in these correlations are scored in opposite directions. The obtained non-significant correlations suggest that the relationship between anality and political aggression is dependent on keying items in the same direction.

These results clearly suggest that a response set to acquiesce, not the reported anality-aggression relationship, adequately accounts for Farber's findings. A possible objection can be raised, however, to the scales developed by the author as inadequate measures of anality and political aggression. The low correlations between A_1 and A_2 , and P_1 and P_2 indicate that Farber's scales and those developed in this study are not measuring the same attributes. The most obvious difference between the respective scales is that acquiescent tendencies are weighted differently in Farber's and the author's tests. The reliabilities reported in Table 1 indicate that A_2 and A_1 are about equally homogeneous. The author's test of anality is thus as internally consistent as is Farber's scale. However, since P_2 appears less reliable than P_1 , Farber's test of political aggression seems the more adequate measure. Nevertheless, the total scales when scored for anality and political aggression are fairly

homogeneous, as evidenced by their Kuder-Richardson reliability coefficients (Formula #20):

$$A_T \dots r = .53$$

$$P_T \dots r = .47$$

These reliabilities are high enough to permit a correlation of $+.50$ between A_T and P_T . The failure of the obtained correlation to approach this maximum cannot be attributed to unreliability of A_T and P_T due to the items added in this study.

REFERENCES

1. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.
2. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475-494.
3. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3-31.
4. FARBER, M. L. The anal character and political aggression. *J. abnorm. soc. Psychol.*, 1955, **51**, 486-489.
5. LORGE, I. Gen-like: halo or reality. *Psychol. Bull.*, 1937, **34**, 545-546.
6. SHELLEY, H. P. Response set and the California attitude scales. *Educ. psychol. Measmt*, 1955, **16**, 63-67.

Received April 26, 1956

A FIRST APPROXIMATION TO NATURE-NURTURE RATIOS FOR ELEVEN PRIMARY PERSONALITY FACTORS IN OBJECTIVE TESTS^{1,2}

RAYMOND B. CATTELL, GLEN F. STICE, AND NORTON F. KRISTY

Laboratory of Personality Assessment and Group Behavior, University of Illinois

No argument is necessary to convince the scientific psychologist that personality theory, clinical psychology, and learning theory are bereft of knowledge of about half the influences at work, in the absence of an understanding of heredity. All learning and adjustment is limited by inherent properties of the organism. Only the unintelligent or uninformed clinician or educator practices individual guidance without estimating what degree of difficulty a given adjustment or performance offers to a particular personality. The general laws of psychology, explaining the genesis of abilities, temperament, and dynamic traits, need also to be more concerned with maturational trends, innate dispositional rigidities, and the effects due to hereditary limits.

Our particular concern is personality theory. The development of an objectively-based personality theory has required: (a) determination of personality structures and processes—a taxonomy of behavior—in which, so far, the factor-analytic and related multivariate methods have been the chief tools; (b) an attempt to penetrate the meaning and

origin of the general and dynamic factors thus discovered, through poststructural experiment. The present study belongs to this second "interpretive" phase of such a long-term research program. For, in the design of an experiment to determine the origins, developmental conditions, and associated criteria of such structures, nature-nurture investigation constitutes a vital first step. Indeed, it is a deficient scientific training which leads psychologists to spend much research time on specific learning or physiological hypotheses, each implying largely either a social mold or genetic influence, without first checking whether nature or nurture is the main source of the structure in question.

An economic and potent research strategy would first halve the possibilities of erroneous hypotheses by undertaking nature-nurture variance ratio determinations on each personality dimension, immediately after its establishment as a replicatable, stable, factor pattern. Admittedly the demonstration that a certain factor is, say, wholly environmentally determined, does not tell us anything about the *specific cause*—intrauterine, infantile experience, sociological institutional impress, etc. But it starts us looking for the right *kind* of quarry.

¹ This research was supported in major part by a research grant from the Institutes of Mental Health, of the National Institutes of Health, United States Public Health Service.

² Besides help from Dr. C. Rao, mentioned in the text, the writers owe a debt of gratitude to Dr. Duncan Blewett, who worked on the project in its later stages, to Drs. W. Kappauf and W. Madow for comments on statistical procedures, and Dr. Touchberry and N. Morton for advice in the field of genetics. They are greatly indebted to helpers in three cities, without whose unquestioning generosity so widely integrated an enterprise would not have been possible. In particular they wish to express their thanks to Dr. Benjamin Burack and George Hartman for facilities at Roosevelt College; Dr. Stuart Cook for laboratory space at New York University; to Dr. Kenneth Lund of the Chicago Schools Psychological Service, for help in finding subjects and Dr. Robert White for facilities at the Harvard Psychological Clinic. They are indebted to Misses Evans, Mourier, and Stewart for generating good motivation in the children in the test sessions and to several social agencies for providing access to their cases.

The expression "wholly environmental" may startle some readers, since it is customary to assert dogmatically that *all* behavior is partly environmental and partly hereditary in its determination. This is true; but a factor is not behavior: it is an abstraction from behavior. And we need to keep an open mind—until more is known about factor dimensions—for the possibility that the influence in covariation which we call a factor could itself be a purely hereditary or environmental influence. However, the balance of information at present suggests that most first-order factors, e.g., the primary abilities, do *not* lie at this particular causative depth. Their unitariness is that of an operating, functional unity in behavior, and as such

they admit of assigning so much of their individual variance to hereditary and so much to environmental variations.

Unfortunately, recognition of the above primary strategy by leading investigators has not resulted in a determined attack to reduce the methodological confusion and conceptual pitfalls (13) hitherto discouraging students from research in psychological genetics. Besides, the general impotence of previously available personality tests, witnessed by the vast accumulations of negative findings in *all* the fields of their application, has basically discouraged hopes of significant results on personality in this field.

The present research makes two innovations:

1. The introduction of *structured personality measurement* in terms of established factors and by means of objective (i.e., behavioral, not questionnaire, self-evaluative) batteries.
2. Development of the *multiple variance analysis method* in place of traditional twin comparison techniques.

THE ELEVEN PERSONALITY SOURCE TRAITS

The theoretical setting and historical evidence for primary personality factors have been sufficiently described in various earlier publications (4, 5, 9, 17, 18, 19, 23) but may be briefly recapitulated. Personality structure analysis, by R and P techniques, has been carried out in three media: (a) life record behavior *in situ* (L data), (b) questionnaire responses (Q data), and (c) objective test behavior (T data); each by cross sections at several age levels. The factors have shown persistence across life-record and questionnaire fields, but patterns in these media have not projected themselves so clearly or uniformly into the objective test realm. Accordingly nature-nurture ratios for eleven L- and Q-data factors have been examined and reported separately (15), without assumptions that they match factors in the objective tests, though in a few cases there is evidence that the same dimension is expressed through all three media.

Some 12 to 14 factors in objective tests have shown satisfactory persistence from middle childhood through adult life, and recent unpublished research shows most of

them to exist in embryo also at 4 and through 7 years. The present research is concerned with the best established 10 factors in the 9-through 14-year range, all of which appear at the adult level also. In addition an eleventh factor (since abilities are part of personality) is included, namely, general intelligence. It was the confirmation of the existence and possible measurement of the adult factors at the 9-14-year level which made the present research feasible (19), for the practical difficulties of demanding so much testing time from so many adults would have exceeded our resources.

The choice of factors is thus made primarily on the basis of their clarity of definition and their stability through childhood and adult life. Their appearance in coordinated studies across three media also assures them a general psychological importance. Moreover, there exist already certain promising hypotheses about their natures (12), which permit referring our findings to clinical and other fields of application.

However, to assert that our researches have established clarity of factor pattern and stable replication of pattern in cross validations is not to claim that these personality factors can yet be measured with the substantial saturation (internal validity) of, say, general intelligence. The reliabilities, and the internal validities in terms of factor estimate, remain woefully below those of old established factors, and are likely to do so until various laboratories concentrate on particular factors. Such research concentration is necessary to find and refine new tests of higher factor saturation, in the light of further hypotheses about the nature of the functions now revealed. The justifications for attacking the problem without waiting a decade for such improvements are that: (a) even approximate answers would lighten the total darkness of present ignorance about personality inheritance in the normal range. The present practice of proceeding from general impressions of constitutional influence might thus at least be more cautiously checked against some empirical data; (b) such indications are necessary in the first place to put us on the track of correct hypotheses by which the choice of better measures for the factors can be guided; and, (c) one illustration of the new method has more value

in guiding research than ten abstract statements of the design.

The 11 factors measured, and the tests used to measure them, will now be briefly described. Factors are today best indexed primarily by a number—for which we have devised in conjunction with other psychologists a Universal Index³ numbering system (7)—and secondarily by tentative, and, largely, descriptive names. Both reference systems now run uniformly through adult and child factors. Usually a minimum of four tests were taken per factor, and were chosen partly for having the better loadings, in the main previous study with children (19), partly for being varied in nature (to avoid any risk of common specifics occurring in a new setting), and partly on good practical reliability as tests. The theory, administration, and scoring of these tests is given, together with test specimens, in the Objective Analytic Personality Factor Batteries (11).

Universal Index No. 1: General Intelligence

Scale 2b of the Cattell Culture Free Intelligence Test was used. It has four subtests: Series, Classifications, Progressive Matrices, and Topological Conditions, which take 14 minutes of actual testing time and have been shown to have a mean saturation with the general factor of .81.

Universal Index No. 16: Assertiveness

1. High expansiveness of copied design. (M.I.⁴ 247)
 2. Unreflective acceptance of unqualified statements. (M.I. 106)
 3. Fast tempo tapping. (M.I. 268)
 4. Fast arm-shoulder movements. (M.I. 270)
- This factor, in all its known manifestations,

³ U.I. numbers 1 through 15 are assigned to the primary ability factors, as evaluated by French's committee (22) and represented by his test kit. The numbers U.I. 16 through 33 were then assigned to the 18 personality factors, in diminishing order of variance, and are represented by the test kit for the O-A batteries (11). The descriptive titles are those settled upon in (12).

⁴ M.I. stands for Master Index Number, by which the test measure is constantly defined in various researches (9, 11, 17, 18, 19) and, in particular, in the Personality Test Encyclopedia in this laboratory, as well as in the Objective-Analytic Personality Test Battery Handbook (11).

is one of general competence and assertiveness, though independent of intelligence. It covers success and aggressiveness in exploration, high fluency, social taste, the C.M.S. test, etc., and is perhaps the same as the "need for achievement" studied by McClelland (26). It is significantly positively associated with success in flying (16).

Universal Index No. 17: Inhibition

1. Tendency to perceive threatening objects. (M.I. 105)
2. Absence of questionable preferences. (M.I. 21)
3. Willingness to play practical jokes. (M.I. 218)
4. High level of experience. (M.I. 197)

This factor, in its full test manifestation, consistently represents cautious, self-critical and timid behavior. It may align with Eysenck's dysthymia factor (20) or with schizothyme tendencies.

Universal Index No. 19: Critical Practicality

1. High index of carefulness. (M.I. 51)
2. High ratio color to form bias in sorting. (M.I. 39)
3. High activity in blind stylus maze. (M.I. 82)
4. High verbal-numerical ratio on Stanford Achievement. (M.I. 99)

A source trait of conscientious exactness and reliability in a wide range of performances (12).

Universal Index No. 20: Comention

(Here actually scored in reverse direction, i.e., the opposite direction to that directly indicated on all tests below.)

1. Honesty in admitting common frailties. (M.I. 219)
2. Respect for authority. (M.I. 246)
3. Agreement with sound platitudes. (M.I. 144)
4. Acceptance of reality principle. (M.I. 113)
5. Tendency to agree. (M.I. 152)
6. Maintenance of logically inconsistent attitudes. (M.I. 65)

Interpretations of this factor have varied between "Gregariousness," "Honesty," and "Acceptance of Social and Ethical Values" (12).

Universal Index No. 21: Exuberance

1. Low accuracy in Gestalt Completion. (M.I. 146)
2. Fast speed of judgment. (M.I. 288)
3. Fluency
 - a. On Sentence Completion (M.I. 271)
 - b. On Nonsense Drawing Associations.
4. Maturity of opinion. (M.I. 34)
5. High ratio consonant to dissonant recall. (M.I. 38)

This source trait shows itself generally in good immediate memory, fluency, fast judgment, and manifestations of ego strength. It is an expansion of factors earlier called "fluency of association" and "cyclothymia."

Universal Index No. 22: Corticalertia

1. Ideomotor Speed:
 - a. Fast psychomotor speeds on simple tasks. (M.I. 6)
 - b. Fast psychomotor speeds (speed for writing).
2. High ratio speed to accuracy in open pencil mazes. (M.I. 120)
3. Fast reaction time. (M.I. 4)
4. Low variance in simple reaction time. (M.I. 263)

The essence of this factor is speed in basic, simple nervous responses, resembling hyperthyroidism or Halstead's (25) "Power" factor. It has the highest prediction of any personality factor of pilot success (16).

Universal Index No. 23: Neural Reserves-vs.-Neuroticism

1. High motor rigidity. (M.I. 2)
2. Low level of aspiration in coding. (M.I. 148)
3. High ratio color to form preference in pictures. (M.I. 227)
4. High body sway. (M.I. 42)

This is the factor which has been shown by Eysenck's criterion rotation to be the main distinguisher of neurotics and nonneurotics (20), though our interpretation has been that it represents some sort of "reserves-vs.-exhaustion" factor, rather than the conflict itself.

Universal Index No. 26: Self Sentiment Control

1. High general annoyability. (M.I. 211)
2. Low ratio annoyance to people and things. (M.I. 242)

3. High memory for attitude statements. (M.I. 167)
4. Few erroneous reactions in complex reaction time. (M.I. 71)

This pattern is associated with strength of will, and self-examination. After intelligence and U.I. 24 it appears the most important factor in determining school achievement.

Universal Index No. 28: Asthenia

1. Amount considered possible in given time for self. (M.I. 191)
2. Low attainability of distant goals. (M.I. 212)
3. High shift toward successfals. (M.I. 193)
4. High shift toward authority. (M.I. 35)

This subtle pattern combines sociability and emotionality with some evasiveness of reality. It has complaisant, "social climber" qualities, but for reasons given elsewhere (12) is finally given the above title.

Universal Index No. 29: Immediate Over-responsiveness

1. High motor rigidity. (M.I. 2)
2. High severity of judgment. (M.I. 133)
3. Disagreement with majority. (M.I. 178)
4. High ratio regular to irregular reaction time. (M.I. 5)

Like U.I. 20 this happened to be scored on the original tests in the opposite direction from the pole to which the title is assigned, but, of course, this has no effect on the variance relationships. The dimension represents a tendency to be highly responsive and tense in relation to social demands. Basal metabolic rate runs high (low to agree with above test labels) and there is probably relation to the Tension factor, Q_4 (10).

THE MULTIPLE HYPOTHETICAL VARIANCE ANALYSIS METHOD

The well-known defects of the twin approach to nature-nurture ratios, to which, for lack of better method, psychologists have perhaps been too inclined to resign themselves, are: (a) too great dependence on our dubious ability completely to separate identicals from fraternal; (b) a systematic danger of unfairly rejecting the more similar fraternal from the latter; (c) suspicion that the environmental differences within fraternal pairs may

be systematically greater, or at least different, from those of identicals; (d) an experimental outcome in the form of a ratio specific to twins and within family conditions, and thus not widely applicable to major theoretical and practical problems. So categorical is the restriction from the last that most investigators really use the approach (as close inspection will show) only to *rank* traits relative to nature-nurture effects, rather than to obtain the ratios which are of basic interest for social and clinical application.

The multiple hypothetical variance analysis method proposed by the senior author (8) and exemplified so far only on factors measured by the Junior Personality Quiz (14, 15) attempts a more general solution than the twin method. It designates four sources of individual differences in any trait: *between-family environmental* differences; *between-family hereditary* differences; *within-family environmental* differences and *within-family hereditary* differences. Its primary advantage is that these variances, and their interactions, are what actually need to be taken into account in the predictions of the clinician or educational psychologist, as well as in theories of personality formation.

Since this paper is primarily for psychologists rather than geneticists, the reader wishing to pursue a fuller examination of assumptions and further scope is urged to read the purely methodological discussion elsewhere (8, 13, as well as 15). Essentially one seeks a number of experimentally ascertainable variances, e.g., between siblings reared apart, unrelated children reared together, etc., which combine the above four sources of variances in different ways. By juggling a number of simultaneous equations, each thus equating a combination of the "abstracted" or "causal" variances to an empirically obtained numerical value, one can hope eventually to solve for these unknown abstracted variances.

For example, the ordinary within-family variance for siblings reared together can be written:-

$$\sigma^2_{ST} = \sigma^2_{wh} + \sigma^2_{we} + 2r_{wh,we}\sigma_{wh}\sigma_{we} \quad [1]$$

where *wh* = within-family hereditary, and *we* = within-family environment sources. That is to say, the variance is the sum of the two possible sources of variance, plus the

covariance, if any. Similarly, for siblings reared apart:

$$\sigma^2_{SA} = \sigma^2_{we} + \sigma^2_{wh} + \sigma^2_{be} + 2r_{wh,we}\sigma_{wh}\sigma_{we} \quad [2]$$

Subtracting 1 from 2 we obtain

$$\sigma^2_{be} = \sigma^2_{SA} - \sigma^2_{ST}.$$

It has been shown in the fuller examination of the multiple variance method⁵ (13, 15) that some fifteen or sixteen equations can be set up, which might seem sufficient to solve for four unknown, abstracted variances, and the six correlations defining possible covariances. But this easy solution is not possible because (a) the cost of obtaining certain groups, e.g., identical twins reared apart from birth, half-sibs, etc., was too great for our field-work resources, (b) the simultaneous equations are always quadratics, whence indeterminate solutions multiply, (c) several equations turn out to be simple linear combinations of others, and so cannot be used; (d) there are actually more than ten unknowns, if we agree that within-family environmental variances as between identical twins, fraternal twins, and ordinary sibs are three different things.

We were able to get data for the following within-family variances, in each case calculated as $\frac{\sum d^2}{k}$ where there are *k* families (pairs of children) and *d* is the deviation of each individual from the mean of the pair. (There are thus *2k* individuals and *k* degrees of freedom.) (a) Identical twins, 104, (b) Fraternal twins, 30, (c) Sibs raised together, 164, (d) Sibs raised apart from birth, 60, (e) Unrelated children raised together both in a foster family from birth, 82, (f) Random pairs (unrelated and apart) from the general population, 210. With the two equations already set out above, and the four following, six become available.

$$\sigma^2_{ITT} = \sigma^2_{we} \quad [3]$$

⁵ Note that this method differs from the ordinary analysis of variance familiar to most psychologists in two important respects: (a) that it deals with abstracted variances, not literal within and between variances corresponding to controlled sources of variance; (b) its goal is an estimate of the actual value of variances, not their use to determine confidence limits for a hypothesis.

$$\sigma^2_{FTT} = \sigma^2_{wh} + \sigma^2_{we} + 2r_{wh,we}\sigma_{wh}\sigma_{we}$$
$$\sigma^2_{UT} = \sigma^2_{wh} + \sigma^2_{we} + \sigma^2_{bh} + 2r_{wh,we}\sigma_{wh}\sigma_{we} + 2r_{we,bh}\sigma_{we}\sigma_{bh}$$
$$\sigma^2_{UA} = \sigma^2_{wh} + \sigma^2_{we} + \sigma^2_{bh} + \sigma^2_{be} + 2r_{we,wh}\sigma_{we}\sigma_{wh} + 2r_{be,bh}\sigma_{be}\sigma_{bh}$$

[4]
[5]
[6]

where *ITT* is identical twins together; *FTT*, fraternal together; *UT* unrelateds together, and *UA*, unrelateds apart (general population).

However, all six cannot be used, because of linear dependence. In fact we can use either of two alternative sets of five. Set I, which used Equations 1, 2, 3, 5 and 6, and was used in the questionnaire⁶ study (15), solves for no correlation, but allows fraternal twin environmental variance, $\sigma^2_{we'}$ to differ from σ^2_{we} , as follows:

$$\sigma_{we} = \sqrt{\sigma^2_{ITT}}$$
$$\sigma_{be} = -r_{be,bh}\sqrt{\sigma^2_{UT} - \sigma^2_{ST}}$$
$$\pm \sqrt{\sigma^2_{UA} - r_{be,bh}\sigma^2_{ST} - (1 - r_{be,bh})\sigma^2_{UT} + 2r_{we,wh}\sigma_{ITT}\sigma_{wh}}$$
$$\sigma_{wh} = -r_{we,wh}\sqrt{\sigma^2_{ITT}}$$
$$\pm \sqrt{\sigma^2_{ST} - (1 - r^2_{we,wh})\sigma^2_{ITT}}$$
$$\sigma_{bh} = \sqrt{\sigma^2_{UT} - \sigma^2_{ST}}$$
$$\sigma_{we'} = -r_{we,wh}\sigma_{wh}$$
$$\pm \sqrt{\sigma^2_{FTT} - \sigma^2_{wh}(1 - r^2_{we,wh})}$$

[7]
[8]
[9]
[10]
[11]

Set 2, using 1, 3, 4, 5, and 6, allows only one kind of within-family environmental variance but solves for one correlation instead, as follows:

$$\sigma_{we} = \sqrt{\sigma^2_{ITT}}$$
$$\sigma_{be} = \sqrt{\sigma^2_{SA} - \sigma^2_{ST}}$$
$$\sigma_{wh} = -r_{we,wh}\sqrt{\sigma^2_{ITT}}$$
$$\pm \sqrt{\sigma^2_{ST} + (r^2_{we,wh} - 1)\sigma^2_{ITT}}$$
$$\sigma_{be} = \sqrt{\sigma^2_{UT} - \sigma^2_{ST}}$$
$$r_{be,bh} = \frac{\sigma^2_{UA} - \sigma^2_{UT} - \sigma^2_{SA} + \sigma^2_{ST}}{2\sqrt{\sigma^2_{SA} - \sigma^2_{ST}}\sqrt{\sigma^2_{UT} - \sigma^2_{ST}}}$$

[12]
[13]
[14]
[15]
[16]

Both sets were used here, so that some direct information on *r* might be obtained

⁶ Note that in Equation 5 a misprint in (15), page 135, put $-r_{we,wh}$ where $-r_{we,wh}\sigma_{wh}$ should have appeared. But the correct solution was used in computation.

from the second; but in terms of available data Set 2 is slightly preferable, because of the weakness referred to in fraternal-twin data.

Now to base solutions on five equations some fairly complex, interrelated assumptions have to be made, which are discussed below, under the title of the *Internal Consistency Method*. Parenthetically, these assumptions are not as gross as in previous work, which has invariably ignored the problem of correlated variances completely. Since there are two *rs*— $r_{be,bh}$ and $r_{we,wh}$ —which cannot be assumed zero, the internal consistency procedure is to try out all possible values (or, at least, at .10 intervals from -1.0 to $+1.0$) and settle on the values which give greatest internal consistency among the values which depend upon them and those which do not. These consistencies are:

1. That if environmental and hereditary influences prove to be correlated, and with a certain sign, in the directly calculable *between-family r*, the unknown *within-family r* is likely to be of the same general nature.

2. That the environmental variance between sibs is likely to be of the same order, probably slightly larger (not ten times larger or smaller), than that between twins: i.e., that σ^2_{we} should be equal to or be greater than $\sigma^2_{we'}$, and $\sigma^2_{we'}$.

3. That if environment is more important than heredity between families, i.e., if $\sigma^2_{be} > \sigma^2_{bh}$, the same general order is likely to exist within families, i.e., $\sigma^2_{we} > \sigma^2_{wh}$.

4. That σ^2_{wh} , which depends on an unknown *r*, should tend to be of the same order, as σ^2_{bh} , which can be determined directly, i.e., *without* assumption. It has been shown elsewhere (13) that with random mating, no dominance or epistacy, σ^2_{bh} and σ^2_{wh} are exactly equal. Complete dominance, in a diallelic model, lowers the ratio to about 0.6, while assortative mating of the order of about $r = 0.5$ raises it almost equally in the opposite direction. Accordingly, out of all the possible values of σ^2_{wh} yielded by assuming different *rs* only a narrow range is acceptable, and therefore the unknown *r* is determined within a narrow range.

The above four internal consistencies are intended both to fix, within limits, the assumable $r_{we,wh}$; and also, to reject one of two alternative quadratic solutions, when one is not

already negative or imaginary. They are only probabilities, but if they agree in converging on a narrow range of values, its probability, as the product of these probabilities, is high. In fact, as shown below they *do* tend to converge, i.e., there is a marked tendency for the $r_{w.e.wh}$ value required by one to be that also required by the others. Also, most quadratic alternatives vanish of themselves by reason of negative or imaginary standard deviations.

These principles of internal consistency and the values which they affect, will become clearer in the actual illustrations in the section after next.

THE EXPERIMENTAL PROCEDURE

The children for the present research were ten- to fifteen-year-olds gathered through the school systems and welfare and placement agencies of Boston, Chicago, and New York. Unrelated children reared together were selected for being reared together from the first year of life. Due to an advertisement of our research as concerned with identical twins the proportion of fraternal twins turned out to be less than expected; this was corrected for. Otherwise there were no obvious selection effects, except the possibility that unrelated children reared together, reached through adoption agencies, are of lower social status than the general child population directly contacted in the schools. Great effort proved necessary to get the SA and UT cases in sufficient numbers. Only same-sex pairs were accepted, to give necessary comparability with the identical-twin group, though both male and female pairs were taken, in order to give greatest generality of application to our subsequent findings.

Each child was brought to the laboratory for testing and undertook four and a half hours of

testing in two sessions, with suitable standard rest pauses and rewards. At the same time, in the case of twins, the physical measurements, fingerprints, iris photographs, taste reactions, etc., were made on which subsequent division into fraternal and identicals was made. In principle this separation was based on the method of Muller (27).

In scoring, the raw scores for each subtest (designated in Section 2 above) in a given factor battery were brought to approximately the same standard deviation (actually scatter, by interquartile range) before being added, to give equal weight to all subtests. The reliability of each factor score was determined as the correlation between two halves of the battery. These halves contained different kinds of subtests (usually two in each). For more exact determination, the coefficient was calculated as the mean of three coefficients—tests (1 + 2)-vs.-(3 + 4); (1 + 3)-vs.-(2 + 4); and (1 + 4)-vs.-(2 + 3)—and gave the values (corrected to battery length) for the eleven factors as shown in Table 1.

These values in some cases run low, but this is the price of working in a field of newly established factors, and of working with independent dimensions instead of conglomerates. Subsequent results, as far as Factors 16, 19, 26, and 28 are concerned, must in consequence be regarded as mere indications. On the other hand it should be realized that our method of computing yields a *factor reliability*, which is decidedly lower than the actual *test battery* reliability, such as would result from a symmetrical split of each and every test in the battery, because it omits correlation due to specific factors. Thus it may give the unwary a poorer impression of the reliability of our testing per se (which was as high as tests of this length normally achieve); but it is this *reliability of measurement of the estimated factor*

TABLE 1
RELIABILITY COEFFICIENTS FOR FACTOR MEASUREMENTS ON ALL GROUPS

Experimental Group	Number of Cases	Universal Index Identification Number for Factors											Symbol in Equations
		1	16	17	19	20	21	22	23	26	28	29	
Id. Twins Tog.	104	.75	.35	.31	.18	.70	.48	.61	.16	.18	.21	.32	σ^2_{ITT}
Frat. Tw. Tog.	30	.70	.32	.35	.14	.58	.40	.54	.31	.10	.18	.32	σ^2_{FTT}
Sibs Tog.	164	.88	.18	.38	.21	.63	.41	.62	.16	.14	.19	.39	σ^2_{ST}
Sibs Apart	60	.84	.10	.52	.25	.63	.39	.58	.38	.29	.31	.46	σ^2_{SA}
Unrel. Tog.	79	.80	.10	.39	.18	.77	.49	.55	.22	.14	.24	.41	σ^2_{UT}
Gen. Pop.	210	.82	.22	.45	.26	.74	.55	.59	.43	.19	.27	.45	σ^2_{UA} (or GP)

behind the tests which is properly to be used in our subsequent calculations.

Finally, an age correction was made to bring all children's scores to the same reference point, namely 12.0 years. This is necessary in *STs* and *UTs* because otherwise their differences would be incomparable with *ITTs* and *FTTs*, for age differences would be added to the true differences. It is necessary to correct to 12.0 years, even in propositi of the same age, because later work on *between-family* means variances would otherwise be affected, and because the magnitude, e.g., of a twin difference, may increase with age. In the case of the intelligence factor, where we already had standardization data, the situation was handled simply by conversion of raw score to IQ. It is noteworthy that most of the personality-factor total scores showed distinctive age trends, up or down, as steady as those for the intelligence score. The plots are set out elsewhere as a supplement to the O-A Handbook (11) and in (12).

Details of calculation of variance from factor scores are as stated above and set out previously for questionnaire measures (15). However, it may be pointed out that where two-member and three-member family results were combined, as in the *ST* and *UT* groups, the

variance was calculated as $\frac{\Sigma x_k^2 + \Sigma x_l^2}{k + 2l}$ where

k and l are the numbers of two- and three-member families respectively, since there is one degree of freedom in the former and two in the latter. Now a device was tried here, as in the questionnaire results, which seems obviously needed in those groups, e.g., the fraternal twins, which, as we now know, fall distinctly short of the numbers needed to produce the stability-of-variance estimate required. (See discussion (13) on fiducial limits by hypothetical variance estimates). If the simple deviations of individuals from family means are plotted a distribution form well known from various twin studies, e.g., those of Thurstone and Vandenberg, is found, shaped like an inverted soup ladle, i.e., a frequency rising briefly from the zero score and then falling slowly to the high values. A best-fitting curve, with the same mean and sigma, can usually be fitted with confidence to the slightly irregular soup-ladle histogram obtained with samples of

this size. This curve would appear to be a far better basis from which to calculate the variance; for otherwise the irregularities, especially toward the higher end of the curve, introduce quite disproportionate distortion when the deviations are squared. Sheppard's correction was applied to all calculations from histograms with less than 12 categories.

However, a second correction exists which may also be used or rejected according to judgment. As pointed out in the earlier results, the observed variance is the sum of true and error variance (15). Through multiplying each obtained variance by the reliability coefficient for that sample an estimate of the true variance is obtained. (A discussion of the effect of reliability upon the error variance in a difference score is given elsewhere (13).) But the reliability coefficients on small groups are rather unstable, so we have used for each subgroup the mean between the obtained coefficient and that existing for the large group. Providing the samples were big enough to give confidence in the value of the reliability coefficient there is no doubt that this error-estimate-corrected variance is preferable to raw variance, for, as stated at the outset, we are interested in the nature-nurture ratios for the *factors*, not merely for the tests. The present case leaves some doubt because of the instability of the coefficients in small groups. Accordingly, in Table 2 we have set out the two extremes in treatment of the basic data, for illustration. Table 2a contains the raw variances, in the raw scores of the various factor batteries, while 2b contains the variances of increased stability through smoothing of the histograms, and further corrected for factor measurement unreliability. It is on Table 2b that we would place greatest emphasis, in view of what has been said above, but comparisons will be made later among results from all four possible sets of data, i.e., including also smoothed but not measurement-error-corrected, and measurement-error-corrected but not smoothed.⁷

⁷ Materials supplementary to Table 2 have been deposited with the American Documentation Institute. Order Document No. 5096 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Make checks payable to Chief, Photoduplication Service, Library of Congress, remitting \$1.25 for microfilm or \$1.25 for photo-prints.

TABLE 2
VARIANCES, RAW AND CORRECTED, FOR ALL FACTORS ON ALL GROUPS

Experimental Group	Number of Cases	Personality Factor U.I. Number											Symbol in Equations
		1	16	17	19	20	21	22	23	26	28	29	
2(a) Values Derived from Raw Score Histogram													
Id. Twins Tog.	104	242.53	10.50	8.03	7.51	21.09	4.49	8.84	7.60	12.60	12.83	11.05	σ^2_{ITT}
Frat. Tw. Tog.	30	267.31	10.90	10.80	3.80	18.50	12.55	16.66	13.85	13.10	10.50	11.45	σ^2_{FTT}
Sibs Tog.	164	373.85	9.25	11.52	8.86	30.28	17.99	16.26	14.15	8.65	14.55	12.66	σ^2_{ST}
Sibs Apart	60	333.44	12.42	15.97	14.66	28.54	24.35	20.32	18.94	27.19	25.04	17.71	σ^2_{SA}
Unrel. Tog.	79	517.30	8.28	14.37	14.95	52.25	22.33	16.81	15.24	12.25	12.94	9.63	σ^2_{UT}
General Pop.	210	614.94	13.78	17.93	14.92	43.90	29.03	21.80	17.03	14.79	19.67	16.43	σ^2_{UA} (or GP)
2(b) Derived from Best-fitting Smoothed Curve: Reliability Corrected													
Id. Twins Tog.	104	179.10	3.08	2.59	1.29	16.74	5.21	5.75	2.23	2.18	2.55	3.13	σ^2_{ITT}
Frat. Tw. Tog.	30	175.32	3.10	3.38	1.02	10.80	6.98	8.54	3.79	1.20	2.21	4.32	σ^2_{FTT}
Sibs Tog.	164	301.55	1.88	4.55	1.90	16.70	7.51	10.50	1.93	1.45	2.66	4.87	σ^2_{ST}
Sibs Apart	60	246.28	1.04	7.76	3.54	21.30	7.21	11.94	7.68	4.51	6.34	8.22	σ^2_{SA}
Unrel. Tog.	79	384.88	.91	5.54	2.40	37.75	10.43	9.39	2.80	2.02	3.37	4.59	σ^2_{UT}
Gen. Popul.	210	402.61	3.00	6.93	3.53	31.81	12.14	9.65	6.05	2.16	4.10	6.54	σ^2_{UA} (or GP)

RESULTS FOR INDIVIDUAL PERSONALITY FACTORS

A problem in condensed presentation arises here. It is desired to show the magnitude of the changes in the end result arising from: (a) basing conclusions on each of the four different ways (Table 2) of arriving at the best determination of the empirical variance values; (b) using Set I equations as compared with Set II; (c) making alternative selections of quadratic equation solutions, and taking various assumed correlations between heredity and environment, i.e., the effect of the processes of judgment involved in the *internal consistency method*.

It has seemed best to present fairly full results for *one* representative factor, in Table 3 below, and to deposit the remaining ten corresponding tables at the ADI.⁸ Factor U.I. 17 is chosen, as of numerically quite typical and psychological interest, through Eysenck's clinical studies under the title *Hysteria-vs.-Dysthymia*, and through our own studies of anxiety and of general personality inhibition. It will be seen in Table 3, and holds over all corresponding tables, that although the four

choices along the continuum from the most raw to the most refined data (illustrated by the three levels in Table 3) affect the absolute magnitudes of the variance solutions a good deal, they influence the nature-nurture *ratios* relatively little. A total impression of internal consistency slightly favors use of the most refined data, under 2b, but any other choice would nowhere radically change our final conclusions.

As to the second choice above (equation sets), it turns out not to be a choice in most cases. The same values are given for σ^2_{we} , σ^2_{wh} and σ^2_{bh} . For σ^2_{be} , however, Set I gives a single unequivocal answer, whereas Set II gives a range of values according to the assumed $r_{we.wh}$. By accepting the Set I answer in Set II a determination of the $r_{we.wh}$ value would follow; thus in Table 3 a particular value between +.10 and -.10 is indicated. Similarly, the two equation sets can be used in a cooperative manner in regard to $r_{be.bh}$. For Set II determines a single value, -.51, and this can be used in Set I to make the dependent σ_{be} value depend on variations in $r_{we.wh}$ only.

It is the third source of indeterminacy above—that comprising quadratic alternatives and the $r_{we.wh}$ freedom not entirely restricted by the considerations of the previous paragraphs—which causes really substantial alterations in the concluded nature-nurture ratios—as a glance at the middle row in each of the six sections of Table 3 will show. ($\sigma^2_{we}/\sigma^2_{wh}$ could

⁸ Materials supplementary to Table 3 have been deposited with the American Documentation Institute. Order Document No. 5097 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Make checks payable to Chief, Photoduplication Service, Library of Congress, remitting \$1.25 for microfilm or \$1.25 for photoprints.

TABLE 3
ALTERNATIVE VARIANCE SOLUTIONS FOR THE SOURCE TRAIT "INHIBITION"—U.I. 17

Using 1st Set of Equations Assumed value for $r_{we.wh}$									Using 2nd Set of Equations* Assumed value for $r_{we.wh}^{**}$								
Variance Source	+.50	+.30	+.20	+.10	.00	-.10	-.30	-.50	Variance Source	+.20	+.30	+.50	+.10	.00	-.10	-.30	-.50
Data from Table 2(a) Raw Histogram																	
σ_{we}^2				8.03					σ_{we}^2				8.03				
$\sigma_{we'}^2$	13.62	11.83	10.76	9.18	7.31	7.29	6.97	(5.11, 2.31)	r_{bebh}				-.51				
σ_{wh}^2	.86	1.44	1.90	2.59	3.48	4.71	8.41	14.21	σ_{wh}^2	.86	1.44	1.90	2.59	3.49	4.71	8.41	14.21
σ_{be}^2	12.18	15.76	14.98	13.18	12.39	(8.24, 5.48)	—	—	σ_{be}^2				4.45				
σ_{bh}^2				2.85					σ_{bh}^2				2.85				
Data from Raw Histogram. Error Corrected									Data from Smoothed Histogram. Not Error Corrected								
σ_{we}^2				2.49					σ_{we}^2				8.36				
$\sigma_{we'}^2$	2.56	2.02	1.99	1.96	1.89	1.82	1.54	—	r_{bebh}				-.51				
σ_{wh}^2	.62	.96	1.19	1.49	1.89	2.37	3.69	5.62	σ_{wh}^2	.88	1.49	1.99	2.66	3.61	4.88	8.76	14.75
σ_{be}^2	6.60	6.15	5.86	5.43	4.93	2.10	1.72	1.24	σ_{be}^2				2.96				
σ_{bh}^2				1.22					σ_{bh}^2				2.23				
Data from Table 2(b) Smoothed Histogram. Error Corrected																	
σ_{we}^2				2.59					σ_{we}^2				2.59				
$\sigma_{we'}^2$	1.66	1.61	1.59	1.51	1.42	1.28	—	—	r_{bebh}				-.51				
σ_{wh}^2	.66	1.00	1.25	1.56	1.96	2.46	3.84	5.90	σ_{wh}^2	.66	1.00	1.25	1.82	1.96	2.46	3.84	5.90
σ_{be}^2	4.97	4.54	4.20	3.69	3.20	2.19	—	—	σ_{be}^2				3.21				
σ_{bh}^2				.99					σ_{bh}^2				.99				

* Note that row 2 is different for the first and second Set of Equations. It represents what is solved for in one and not in the other, i.e., the alternative of getting r_{bebh} or $\sigma_{we'}^2$.

** For economy of space in the table two values have been omitted, additional to those above 50.

range from about $1/10$ to about unity.) For economy the $r_{we.wh}$ range is shown in Table 3 only from +.50 to -.50, with intervals smaller where a finer choice actually needs to be made. (Incidentally, economy also dictates illustrating only three of four data sources on one equation set, though both cover all four.)

These r s can now be examined by our third—internal consistency—principle. First, by our above criterion that interfraternal twin environmental should not differ radically from intersib σ_{we} , i.e. that $\sigma_{we}^2/\sigma_{we'}^2$ should run about unity or a little higher, highly positive $r_{we.wh}$ values in Set I must be rejected, as giving $\sigma_{we'}^2$ values as large as or larger than the fixed obtained σ_{we}^2 . Though most r values from +.40 through high negative would give acceptable $\sigma_{we'}^2$ values, the $r = +.10$ to $-.10$ range, agreeing with that already independently indicated above (or, at any rate, something between an $r_{we.wh}$ of +.10 and $-.20$), is most acceptable. For this gives a between-twin environmental variance (all three empirical

variances, 2a, Raw histogram: Error corrected, and 2b) about 60 per cent through 80 per cent of the between-sib environmental variance. The second internal consistency check—that of the $\sigma_{wh}^2/\sigma_{bh}^2$ ratio—also agrees. For assuming this will be nearer the 1.5 than the 1.0 limit, on account of known assortativeness of mating in regard to temperament, a value of $r_{we.wh}$ between +.10 and $-.05$ gives the best σ_{wh}^2 value, all six tables considered. The σ_{be}^2 value from Set II (4.45, 2.96, 3.21) suggests in all three tables of Set I an $r_{we.wh}$ of 0 to $-.10$. Thus all three indications suggest an $r_{we.wh}$ of +.10 to $-.10$.

The remaining two internal consistency principles, namely, that the ratio and the correlation of heredity and environment within families should be the same as that between families, must be considered of decidedly lower probability, and given less potency in the decision. As Table 3 shows, environment predominates over heredity between families in Set I, and in Set II for any $r_{we.wh}$ to the positive

side of $-.20$ (approximately) as regards *within*-family variance, it turns out (in both sets) that any value to the positive side of $-.20$ will permit environment to predominate there also.

A good convergence is thus obtained on an $r_{we,wh}$ value within the range $\pm .10$. In this case we are inclined to turn our backs upon the fourth consideration, namely the probability that the within- and between-hereditary-environment correlations should be taken to be similar in value. For most weight must always be given to the high probability of a 1.00 to 1.5 ratio for $\sigma_{wh}^2/\sigma_{bh}^2$, and in this case it indicates the $+.10$ limit of the $\pm .10$ range. Similarity of within and between conditions is, in any case, the principle with lowest probability; for one can and does find instances where the social and psychological mechanisms can be seen to be such as would cause divergences. For example, a better hereditary intelligence will tend to generate a better educational environment *between* families; but *within* families parents might concentrate on giving greater intellectual stimulation to more backward members.

The extent of the problem of alternative quadratic solutions is shown by the two answers gathered in parenthesis in Table 3. Most alternatives are negative sigmas or imaginary. Occasionally both are imaginary, as shown by the dotted lines. Of the alternatives in parenthesis, 2.31 can be rejected as too remote from $\sigma_{we}^2 = 8.03$ but in the other instance both would have to be considered if both were not ruled out by choice of $r_{we,wh}$ value.

With this illustration in detail, on one case, of the possibilities and limitations of solution by the internal consistency method, using the most practicable, six group, application of the multiple variance analysis method, we can turn to a brief examination of the actual results, as accepted for further psychological discussion, regarding the ten remaining particular personality dimensions. The table⁹ for General Intelligence corresponding to Table 3 for Factor I yields no value for $r_{be,bh}$, but, fortunately, independent social-psychological investigation suggests an r of about $+.2$ to $+.3$ between intelligence and educational opportunity (2, 3). Internal consistency, notably the $\sigma_{wh}^2/\sigma_{bh}^2$ ratio, indicates an $r_{we,wh}$ of -10 to 0 . A value thus approaching zero is presumably due to the tendency in a familiar group to

adapt environment demand to native capacity, thus correcting, incompletely, the tendency of the world generally to pressure the dull and let the bright mark time. Accepting these correlations, heredity is about twelve times as important *between* families and twice or more *within* families, where hereditary variance is much less. Incidentally, this higher value for heredity than in our earlier study (15) matches our substitution of culture-free intelligence tests (12) and is also in better agreement with the findings of Burks (1) and Holzinger (31).

Factor U.I. 16, *Assertiveness*, has uncertain results due to low reliability in the battery itself. Using raw variances (Table 1) the best conclusion is an overwhelming predominance of environment. If our hypothesis that this is McClelland's (26) "need for achievement" is correct, this makes sense in terms of a tension level almost wholly determined by social stimulation.

Factor U.I. 17, *Inhibition*, has just been discussed from Table 3. The results point to an environment predominance between $1\frac{1}{2}$ and $2\frac{1}{2}$ within the family and $1\frac{1}{2}$ to $4\frac{1}{2}$ between families, with a substantial negative (-0.5), corrective influence of environment *between* families, as might be expected for a trait of inhibition-vs.-lack of inhibition.

Factor U.I. 19, *Critical Practicality*, again suffers from low battery validity, but good internal consistency is given by an $r_{be,bh}$ of -0.3 , and an $r_{we,wh}$ about the same, as shown in Table 4. This might express a curvilinear relation: environmental pressure principally on those who fall *below* average in critical practicality. The tentative interpretation of this as a super-ego component (11, 12) agrees with the environmental predominance.

No quadratic alternatives, and good consistency (when $r_{we,wh}$ agrees with $r_{be,bh}$ at about $-.5$) is found for *Comention*, U.I. 20. Hereditary variance is four times greater than environmental, between families, and is also greater within. A whole class of mooted interpretations of this factor as "ready adoption of culture," "authoritarian personality," etc., has therefore to be rejected. Search had evidently better proceed in the direction of a temperamental "corrigibility-vs.-incorrigibility" (general rigidity?) or altruism of disposition.

Again, in U.I. 21, *Exuberance*, there is a generally clear solution, but the convergence is

⁹ See Footnote 8.

TABLE 4

WITHIN- AND BETWEEN-FAMILY NATURE AND NURTURE VARIANCES FOR ELEVEN PERSONALITY FACTORS

	Personality Source Traits: Universal Index Numbers										
	1	16*	17	19	20	21	22	23	26	28	29
Value for $r_{be, bh}$	.25	no solution	-.51	-.28	-.53	.23	-.80	-.56	-.98	-.91	no solution
Most consistent value for $r_{we, wh}$	.00	-.85	.10	-.20	-.60	-.10	.50	-.50	-.60	-.30	.50
σ^2_{we}	179.1	10.5	2.6	1.3	16.7	5.2	5.8	2.2	2.2	2.6	3.1
$\sigma^2_{we'} \dagger$	52.9	7.3	1.3	[0.2]‡	15.1†	4.7	4.8	3.9	1.8	2.3	2.7
σ^2_{wh}	122.5	0.06	1.8	1.1	24.0	3.1	1.23	1.6	1.2	1.1	.5
σ^2_{be}	6.7	3.2	3.2	1.6	4.6	.4§	2.3	5.8	3.1	3.7	3.4
σ^2_{bh}	83.3	[0]	1.0	.50	21.1	2.4	.9	1.0	.6	.7	[0]

* This factor is based on Table 1a and Equation Set 1 except σ^2_{be} which is on Equation Set 2. Otherwise all data is Table 2b and Equation Set 2.

† Note when $\sigma^2_{we'}$ is taken from other than 2b it is always reproduced as the same fraction of σ^2_{we} as in the other table.

‡ The only solution on this neighborhood is for $r_{be, bh} = .15$ which is as shown.

§ From 1st equation.

only obtained by accepting a low negative $r_{we, wh}$ to go with the obtained low positive $r_{be, bh}$. Heredity and environment have a roughly equal role, within families, but heredity predominates between families. This trait thus behaves very much like intelligence. Families with a generous hereditary endowment develop an atmosphere, or a status in society, somewhat more favorable to further development of the trait; but for individuals within families this does not hold.

Source trait U.I. 22, *Corticalertia*, has special interest because it gives the highest prediction, among personality factors, of success as an air pilot (16). The obtained value for $r_{be, bh}$ of $-.80$ is unusually high and we also find an illustration of the correlation within families being opposite in sign to that between. Presumably within the family the individual of lower alertness is understood, and treatment is to some extent adjusted to him. But between families, which face the same demands from the external world, there is a decided pressure on the less alert to speed up.

Source trait U.I. 23, like U.I. 22, is one of considerable practical and theoretical interest because, as shown by Eysenck and his co-workers (20) it is the factor most significantly distinguishing neurotics from normals. However, Cattell has interpreted and labeled it "Lack of Neural Reserves" rather than as the neurotic conflict itself (9, 12) which agrees better with the large hereditary influence. In the U.I. 23 equivalent of Table 3 it was a little difficult to find an $r_{we, wh}$ value bringing $\sigma^2_{we'}$ near to σ^2_{we} , but otherwise the internal

consistency was very good. Perhaps in this trait there really is a tendency in the twin situation for individuals to take on divergent, supplementary roles.

The result for "Neural Reserves-vs.-Neuroticism," at the point of greatest internal consistency, is a predominance of environment, very marked as between families, less marked within. This agrees, if similar moderate negative r s are taken, with the result we obtained (15) for questionnaire measures of neuroticism—Factor C (14)—but disagrees with the strong hereditary determination found by Eysenck and Prell (21). The latter could be due to defects of the twin method, or to the use of a much smaller total sample than ours, or to the error of calculating variance from the correlation when identicals and fraternal had very different total variance as individual members of a population. In any case our result, especially in giving so much weight (the 5.8 figure, Table 4) to family atmospheres, agrees with clinicians of experience, while denying the doctrinaire clinical theory which gives no role whatever to heredity. Some addition to our knowledge of neurosis may be claimed in the discovery of marked negative correlations, showing that either the individual concerned, or society, puts those of lesser stress tolerance in less stressful positions, and vice versa.

Although the internal agreements in Factors U.I. 26 and 28 are good, we drop detailed analysis because the reliabilities of the batteries available at the time of testing for these factors turned out to be poor (Table 1). In both, and in both settings (but especially

between families) environment is roughly two to four times as influential as heredity, which agrees with the concept of U.I. 26 as an educationally acquired will-control ("Self Sentiment Control") and of U.I. 28 as an "over-educated" asthenia. The extreme size of the negative r s invites checking in later experiments.

On the other hand, the firm, internally consistent finding that U.I. 29, which had tentatively been called "Immediate Over-Responsiveness," has the most complete environmental determination of all factors, demands revision of the older "physiological" direction of interpretation. Presumably it is better conceived as an eager responsiveness largely built up by a favorable environment, and in which signs of such responsiveness are the signal for increased social stimulation, as evidenced by the positive $r_{we,wh}$.

DISCUSSION

Apart from the implications of the obtained results for research on the interpretation of particular personality dimensions and for long-range predictions with particular batteries in applied psychology, there are one or two general emergents needing discussion.

First, it is noticeable that out of eleven factors three show a within-family environmental influence less than the between, three show them about equal and five show a decided within-family preponderance. Thus the results favor a psychologist's rather than a sociologist's view of the importance of environment. Where basic personality is concerned, it is the within-family happenings, rather than differences of social status, etc., which exercise the truly powerful influences. But psychologists themselves may need to be reminded that they should not think of environmental variance as "psychological variance" in environment only. Psychological traits can be influenced by physical, biological, and prenatal environmental variance, e.g., as in Mongolian imbecility, and our figures are for "total environmental variance."

Secondly, there is an unmistakable pre-dominance (14 out of 19) of negative correlations between heredity and environment (and the negative r s average higher). Naturally, one cannot be sure whether this is a linear relation over the whole range or a curvilinear relation

in which heredity and environment are opposed only over certain parts of the trait range; but it implies that hereditary deviants from the mean, at either end or both, are constrained by environmental influences *toward* the mean. Now from a quite different field of experiment—ego defenses and misperception measurements—a law of proportionality of overt and covert dynamic trait strength has been formulated (6). It is a descriptive law, stating that a person deviating in one direction from the mean in an overt trait deviates in the same direction, proportionally, in the amount of covert or repressed tendency for the same trait. The present observation suggests a beautifully supplementary law, of coercion to the biocultural norm. According to this, an individual who deviates from his community norm, set by the biological and cultural central tendencies of his group, will experience a cultural, "educational," constraint toward this mean. Within society it would account for the negative correlations observed here, and within the individual it would account for the proportionality of repressed to expressed trait observed elsewhere (6).

A subhypothesis requires investigation of the significance, in another study, of the indication here that the negative correlation within the family is generally less than that within society. This *could* be a purely statistical effect of differences in magnitude of range, but the within and between ranges we get do not point to any gross influence of this kind. Instead, one may hypothesize 1, that the mere family biosocial norm is not so explicit and institutionalized as the cultural ideal of society, and/or 2, that the small family group (a) can afford to, or (b) is affectionately disposed to, or (c) is better equipped by knowledge of individuals to, tolerate and accommodate to hereditary deviations.

The above law of coercion to the biocultural norm presumably operates in most temperamental and dispositional traits in which, over centuries, culture and biological endowment have accommodated to produce a cultural norm central in the biological range. For example, we have a notion of an appropriate average degree of dominance or independence, which causes us to quell the too-dominant child and encourage the downtrodden. But, perhaps less frequently, one will encounter two

other patterns of biological distribution and cultural ideal: (a) coercion to an ideal at the extreme of the biological and perhaps also of the cultural (overt behavior) range, as in the difference between our sex ideal and the Kinsey data, and (b) what we may call environmental autogenesis. The former pattern should produce zero correlation if the coercion is uniform for all levels of behavior, and a moderate negative correlation if it lessens as persons approach the extreme of attainment of the ideal. The values for *U.I. 19*, which we have considered a super-ego strength factor, appropriately (and alone) fit this pattern.

By environmental autogenesis we mean the tendency of a genetic trait of its own nature to bring about, for an individual or family, an environment favorable to its further development. Since it is hard to imagine a tendency on the part of *society* to encourage deviants from the biological norm in *either* direction (except in a self-destructive society) an autogenic tendency within the individual or family would seem the only explanation for positive correlations. In our experiment a positive $r_{be.bh}$ seems indicated only for intelligence and the temperament trait of fluency and rapid decision called Exuberance (12). It is not difficult to imagine that both native intelligence and native mental energy tend to create, or get into, environments which make still greater demands on these same qualities.

If, finally, we examine these findings, not in regard to their psychological indications for particular traits or general laws, but for the light they throw on the method itself, some broad pragmatic conclusions are perhaps possible. For traits on which previous data exists—intelligence and neural reserves (*U.I. I* and 23)—there is either good general agreement or explicable difference of emphasis. Where only previous “hunches” exist, the difference most commented on by psychologists is the greater role which the present method seems to assign to *within-family* variance of *both* kinds. It is hard to find, however, any firm basis for the view to which our results come as a surprise.

On the other hand, there may be better reason for suspecting some tendency in the method toward a slight overemphasis of environment as a whole, for both in this and the previous research (15) environment is the preponderant influence in a majority of the per-

sonality dimensions. (The mean of the percentage contributions for eleven factors is σ_{we}^2 : 39.8%; σ_{wh}^2 : 18.7%; σ_{be}^2 : 29.4%; and σ_{bh}^2 : 12.1%, making total environment 69.2% and total heredity 30.8% of the whole.) It may be argued that such a shift of all ratios toward environment would result from test unreliability, whether through experimental error or through diurnal function fluctuation. But our method of subtracting variance for unreliability of factor measurement would eliminate the former. However, the use of the symmetrical consistency coefficient instead of the (unknown) stability coefficient would still *leave function fluctuation in the environmental variance*. Since function fluctuation may be considerable in personality traits (12), this may well be the explanation of our bias. However, the impression of an overemphasis on environment may partly imply comparison with the nature-nurture ratios to which we are accustomed in physical traits, and we should bear in mind that behavior is in general more plastic and labile than physical structure.

Strictly limiting consideration to comparisons with earlier *psychological* results, however, the possibility must be entertained that our taking the environmental difference for sibs rather than twins in our final statement accounts for the change of emphasis from earlier studies, almost exclusively on twins. Sibs, being born years apart, are subjected to the influence of parents of different age and experience, perhaps different gestation and early climatic conditions, and belong in truth to different historical periods. We have corrected sibs to the same age in terms of the average trend in the personality dimension with age, but this should still leave environmental difference greater than that existing within a family at a given moment. It should be noted, incidentally, that this argument would confirm our methodological innovation of separate variance terms for between-sib and between-twin environmental differences. Indeed, direct inspection of our values (Table 4) shows the intersib variance generally to run larger. (The mean of the percentage contributions for the eleven factors is σ_{we}^2 = 39.8%, 29.7.) This result is only to a slight degree an artifact from choice of $r_{we.wh}$, for its choice was dictated by two or three other considerations than getting comparable twin and sib values.

Thus, in general, this first trial of the method may be said to confirm the theoretical expectations of its effectiveness and appropriateness. The one major defect which visibly persists, in spite of valuable comment from Rao (29, and private communication) and some other statisticians, is our inability yet to define confidence limits for hypothesis testing, i.e., to determine standard errors and distributions for the final ratios among the four abstracted variances. In two or three cases it can be done (Equations 12, 13 and 15 above), but elsewhere the abstracted variances are themselves derived as complex functions of several experimentally determined variances, each on a sample of different size.

The possibility of a solution is discussed in a theoretical paper elsewhere (13). But it may be said alike on theoretical grounds (in the few instances where the standard error has been calculable) and from empirical demonstration of instability of results here (according to changes observed from irregular to smoothed histograms, and the observed effects of adding or subtracting a relatively small percentage of our cases) that our 647 sample, though arduously obtained, has proved insufficient to justify regarding the above ratios as more than mere indications. Further research should not be undertaken without resources for field work at least twice as great as ours. Thus our findings are only the first glimmerings in an area of total darkness, and fit at best for guiding hypotheses to be given more exact testing.

Despite this admitted statistical deficiency, it can be said that the method as a whole gets a further pragmatic proof of its effectiveness from the general direction and order of the particular relations indicated being in agreement with the psychological nature of the various factors. Traits which, by their observed psychological character, would be expected to be wholly environmental, prove (with one possible exception) to be so; while the obtained variance correlations fit psychological or sociological theory. For example, the arguments on innate intelligence and environment have been for positive correlation between families and negligible within. The same would be expected for Exuberance (U.I. 19) and the opposite for Inhibition (U.I. 17), which we find.

In conclusion it should perhaps be pointed out for those less familiar with this approach that no theoretical or practical deductions

should assume these ratios to represent absolute values. They are of course entirely relative to the cultural pattern and racial composition of the population studied. Inferences about absolute Mendelian structure are at least one step beyond the ratios, and would require collation of results from differently constituted populations. Parenthetically, and in relation to the above suspicion of too great an environmental emphasis, it may be that dealing with a higher than normal percentage of broken families, as in this design, gives a greater environmental spread than is truly typical of our culture. Comparative cultural considerations of this nature, involving determination of the ratios in relation to *different*, cultural - and - racial - variance - defined populations, should be part of an improved use of the method.

SUMMARY

1. A design of experiment called the multiple-variance analysis method, with certain advantages over the traditional twin-comparison design, and capable of solving for all four variances and their intercorrelations, has been applied to determine nature-nurture ratios for eleven primary personality factors (including intelligence) measured by the Objective-Analytic Test Battery (11).

2. The subjects, each tested for 4½ hours in two sessions, consisted of 104 identical twins, 30 same-sex fraternal, 164 same-sex sibs reared together, 60 same-sex sibs reared apart, and 79 unrelated same-sex pairs reared together, constituting a total of 437, or 647 boys and girls when the "control" general population variance sample is considered. They were aged 10 through 14 years, but all scores were corrected to the same age, 12 years, before calculating variances.

3. The study is entitled an "approximation" because (a) it lacked all the necessary groups to use all equations required for solution of the multiple-variance analysis method and fell back upon "the principle of internal consistency"; (b) the objective test measurement of personality factors needs raising to better levels of reliability and validity; (c) mathematical statistics have not yet worked out confidence limits for these ratios, but results indicate that a minimum sample 2 or 3 times larger than ours would be desirable.

4. The research thus had the double function

of trying out a method and obtaining a first approximation regarding the role of nature and nurture in primary personality factors. Development of the internal consistency modification of the method, which proved reasonably successful, may be important because even when larger samples are obtained conditions may make impractical in many studies the obtaining of data for more than six equations. Discovery of the nature-nurture ratios showed heredity to be the main determiner in two source traits, Intelligence (U.I. 1) and Comen-tion (U.I. 20). Environment largely determined six: Assertiveness (U.I. 16), Corticalertia (U.I. 22), Neural Reserves-vs.-Neuroticism (U.I. 23), Self Sentiment Control (U.I. 26), Asthenia (U.I. 28) and Over-responsiveness (U.I. 29). Heredity and environment were about equally influential in Inhibition (Hys-teria-Dysthymia) (U.I. 17), Critical Prac-ticality (U.I. 19), and Exuberance (U.I. 27).

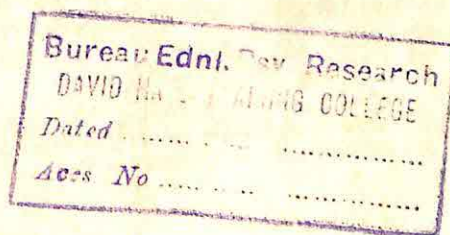
5. The most internally consistent solutions require a majority of negative heredity-environment correlations and a law of *coercion to the biocultural mean* is proposed to account for this. It fits and supplements the law of overt-covert proportionality proposed earlier in the study of trait dynamics.

6. In general, the variance ratios obtained, and the directions of the correlations, fit expectations from the psychological natures of the particular source traits concerned, and are in line with any previously known, independent findings on these traits. This is offered as a broad pragmatic proof that the multiple-variance analysis method, and in particular its specialized development by the principle of internal consistency, is working according to theoretical expectations. However, the method needs the improvements implied under 3 above, notably, (a) field-work resources to use some ten equations and larger samples for each, (b) allowance for sampling selection, and (c) determination of the formulae for confidence limits.

REFERENCES

1. BURKS, B. S. The relative influence of nature and nurture upon mental development. 27th Year-book. *Nat. Soc. Stud. Educ.*, 1928, 219-316.
2. BURT, C. L. *Intelligence and fertility*. London: Hamilton, 1946.
3. CATTELL, R. B. Effects of human fertility trends upon the distribution of intelligence and culture. 39th Yearbk. *Nat. Soc. Stud. Educ.*, Part I, 1940, 221-233.
4. CATTELL, R. B. The description of personality. Principles and findings in a factor analysis. *Amer. J. Psychol.*, 1945, 58, 69-90.
5. CATTELL, R. B. *The description and measurement of personality*. New York: World Book Co., 1946.
6. CATTELL, R. B. Principles of construction of projective or "ego-defense" tests of personality. In H. H. Anderson, *Projective techniques*. New York: Prentice-Hall, 1950.
7. CATTELL, R. B. A universal index for psychological factors. Adv. Publ. No. 3. Urbana: Lab. Pers. Assessment, Univer. of Illinois, 1953.
8. CATTELL, R. B. Research designs in psychological genetics with special reference to the multiple variance method. *Amer. J. Hum. Genet.*, 1953, 5, 76-91.
9. CATTELL, R. B. The principal replicated factors discovered in objective personality tests. *J. Abnorm. Soc. Psychol.*, 1955, 50, 291-314.
10. CATTELL, R. B. Validation and intensification of the Sixteen Personality Factor Questionnaire. *J. clin. Psychol.*, 1956, 12, 205-214.
11. CATTELL, R. B. The objective-analytic personality test batteries, adult and child. Champaign, Ill.: Inst. Pers. Ability Testing, 1604 Coronado Drive, 1955.
12. CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book Co., 1957.
13. CATTELL, R. B. Theory and method in psycho-logical genetics. 1. The chief multiple variance analysis equations. *Psychol. Rev.*, in press.
14. CATTELL, R. B., BELOFF, J. H., & FLINT, D. The junior personality quiz. Champaign, Ill.: Inst. Pers. Ability Testing, 1604 Coronado Dr., 1954.
15. CATTELL, R. B., BLEWETT, D. B., & BELOFF, J. R. The inheritance of personality. A multiple variance analysis determination of approximate nature-nurture ratios for primary personality factors in Q-data. *Amer. J. Hum. Genet.* 1955, 7, 122-146.
16. CATTELL, R. B., CHECOV, L. L., COGAN, E. A., & SCHIFF, H. Psychiatric screening of flying personnel: the prediction of adjustment from an objective test battery. Unpublished Project Rept., Dept. Clin. Psychol., School of Av. Med., U.S.A.F., 1954.
17. CATTELL, R. B., DUBIN, S. S., & SAUNDERS, D. R. Verification of hypothesized factors in one hundred and fifteen objective personality test designs. *Psychometrika*, 1954, 19, 209-230.
18. CATTELL, R. B., DUBIN, S. S., & SAUNDERS, D. R. Personality structure in psychotics by factoriza-tion of objective clinical tests. *J. Ment. Sci.*, 1954, 100, 154-176.
19. CATTELL, R. B., & GRUEN, W. The primary per-sonality factors in eleven year old children by objective tests. *J. Pers.*, 1955, 23, 46-478.
20. EYSENCK, H. J. *The scientific study of personality*. London: Kegan Paul, 1952.

21. EYSENCK, H. J., & PRELL, D. B. The inheritance of neuroticism; an experimental study. *J. ment. Sci.*, 1951, **97**, 441-465.
22. FRENCH, J. W. Selected tests for reference factors (in abilities). Princeton, N. J.: Educ. Testing Service, 1953.
23. FRENCH, J. W. The description of personality measurements in terms of rotated factors. Princeton, N. J.: Educ. Testing Serv., 1953.
24. GUILFORD, J. P., & GUILFORD, R. B. Personality factors D, R, T and A. *J. abnorm. Soc. Psychol.*, 1939, **34**, 1-10.
25. HALSTEAD, W. *Brain and intelligence*. Chicago, Univer. of Chicago Press, 1947.
26. MCCLELLAND, D., & LIBERMAN, A. M. The effect of need for achievement on recognition of need related words. *J. Pers.*, 1949, **18**, 236-247.
27. MULLER, H. J. MENTAL traits and heredity as studied in a case of identical twins reared apart. *J. Hered.*, 1925, **16**, 433-448.
28. NEWMAN, H. H., FREEMAN, F. N., & HOLZINGER, K. J. *Twins: A study of heredity and environment*. Chicago: Univer. of Chicago Press, 1937.
29. RAO, C. R. *Advanced statistical methods in biometric research*. New York: Wiley, 1952.
30. REEVE, E. C. R. The variance of the genetic correlation coefficient. *Biometrics*, 1955, **2**, 357-386.
31. SCHWESINGER, G. C. *Heredity and environment*. New York: Macmillan, 1933.



THE EFFECTS OF RESERPINE ON THE ACQUISITION OF THE CONDITIONED AVOIDANCE RESPONSE¹

ISAAC BEHAR AND ARTHUR J. RIOPELLE

Emory University

A NUMBER of tranquilizing drugs such as the *Rauwolfia* alkaloids and chlorpromazine produce disruption of a well-established conditioned avoidance response (10, 11) which cannot be attributed to impairment of motor coordination. Such drugs are of psychological interest in addition for their potential efficacy in the reduction of anxiety (2, 9), and in the treatment of behavior disorders (4). They are generally called tranquilizers because the subject becomes less fearful and is more manageable, an observation which has also been reported for the monkey (6). Also, the EEG pattern is characteristic of that of a waking individual rather than of a sleeping individual (3, 6, 8).

In order that the therapeutic effectiveness of these drugs be maximal the behavioral alterations should outlast the specific physiological effects. Under this hypothesis, either some anatomical change takes place or a consequence of the experiences undergone while under the influence of the drug survives the physiological adjustment to the drug. For example, the responses to motivational and emotion-provoking situations would be expected to be susceptible to these latter effects. Tranquilized patients thus might be especially responsive to other therapies wherein new adjustments to emotion-provoking stimuli are encouraged. With this in mind, the present study was undertaken to explore the effects of one of these drugs, reserpine,² on the rate of acquisition of conditioned avoidance behavior. Prior investigations of the effects of these drugs on the avoidance response have been confined to animals which had been trained prior to injection.

METHOD

Subjects. Twelve naive *Macacus cynomolgus* monkeys of both sexes, all approximately one year old, were used. Prior to this experiment the animals had been

trained on discrimination problems involving food reward, but had had no experience with the present test apparatus.

Apparatus. The animals were trained in a large wooden shuttle-box divided into two equal compartments by a barrier one foot high running the width of the box. Each compartment was two feet square and five and one-half feet high. The floor of the apparatus consisted of metal rods which could be electrified, current through the bars being held constant at 4.5 ma. delivered in 60 pulses per sec. A 75-watt electric light bulb was located in plain view of the animal beneath the rods of each compartment.

Procedure. After an animal was led into the apparatus, the light bulb was turned on in the compartment which he occupied. Four seconds later, the metal rods in that compartment were electrified for 11 sec., at which time both stimuli terminated. The interval between each trial varied randomly between 10 and 30 sec. Reserpine, when given, was administered (.50 mgm./kgm.) three hours before testing. Previous work in this laboratory had shown that drug effects are nearly maximal at this time. Experience had also shown that effects of the drug on avoidance performance were negligible approximately ten hours after injection. Spacing of sessions a week apart therefore assured an absence of unspecifiable cumulative drug effects.

Experimental design. Subjects were assigned at random to two equal groups. Animals in the first group served as controls and received no reserpine. The procedure for this group consisted simply of three training sessions spaced one week apart. Sixty trials were administered in each session.

The program for the experimental animals was conducted in two phases. During Phase I, the experimental group underwent the identical schedule as the controls but under the influence of the drug. Beginning a week after the last session of testing under reserpine, these animals were given three additional weekly sessions of 60 trials without drugs. This training constituted Phase II.

The intent of the experimental design was to determine whether the 180 trials in which light and shock were paired while the animals were drugged would facilitate or hamper subsequent learning of the avoidance response when the animals were no longer drugged. To accomplish this aim, the learning of nondrugged animals in both groups was compared.

RESULTS

The injected animals exhibited the characteristic signs of reserpine tranquilization, including diarrhea, tremor, and decreased motor activity, but without attendant loss of normal coordinated movement when adequately motivated.

¹ This work was performed under Contract No. AT-(40-1)-1441 with the Atomic Energy Commission.

² Reserpine was kindly supplied by CIBA Pharmaceutical Products, Inc., Summit, N. J.

In the experimental situation an animal can respond in one of three ways: (a) he can *avoid* the shock altogether by climbing over the barrier into the dark compartment within 4 seconds after the presentation of the conditioned stimulus; (b) he can *escape* the shock by climbing the barrier between 4 and 15 seconds after the onset of light; or (c) he can take the shock for the entire 11 seconds (*response failure*). Records were kept of all three response types.

During Phase I, none of the injected animals acquired the appropriate conditioned avoidance response within 180 trials. Indeed, 94 per cent of all recorded responses during this phase were response failures. Five and one-half per cent of the responses were escape responses, while only 0.5 per cent were avoidance responses. No animal made more than two avoidance responses in any 60-trial session.

Fig. 1 illustrates the acquisition data for the control group and for the experimental group during the critical block of 180 trials. The graphs for the control group show the characteristic manner in which monkeys learn the avoidance response in this apparatus. The left panel shows the poor initial performance and the superior terminal performance. All animals attained at least 28 avoidance responses in 30 trials within the training given. The center panel shows that initially the naive animals responded by jumping over the barrier after the onset of the shock. They only infrequently stayed in the lighted compartment for the entire 15 sec. (right panel).

Figure 1 also shows that the experimental group made no more than 50 per cent avoidance responses in any block of trials, even during the last session. The statistical comparison applied to the acquisition data for the two groups was the Alexander test for trend (1). In order to obtain homogeneity of variance and independence of mean and variance, Bartlett's modification of the arc-sin transformation was employed. The difference between group means was significant at the 0.005 level ($F_{1, 10} = 17.9$), but there was no significant difference between group slopes, probably because measures for both groups were anchored to the same value at one end.

The frequency of escape responses and of response failures for the experimental group is especially interesting. Unlike the control ani-

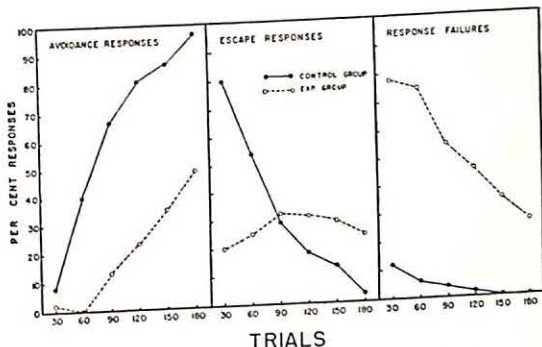


FIG. 1. ACQUISITION OF THE CONDITIONED AVOIDANCE HABIT

A nondrugged control group is compared with an experimental group of monkeys after the effects of reserpine presumably had worn off. Sixty trials were given in each of three sessions a week apart.

mals, which made many escape responses and few response failures, the experimental animals made few escape responses and many response failures, particularly in the early part of training. Even after 180 trials the experimental animals made more than twice as many response failures (27 per cent) than the control animals did initially (12 per cent).

DISCUSSION

The finding that the drugged animals did not jump over the barrier during Phase I accords with the finding of disruption of the response in trained animals. Indeed, the dose level was chosen to insure a high number of response failures.

The more significant finding is the persistent retardation of acquisition in Phase II. This finding is in sharp contrast to the 10-20 hour recovery of the response after reserpine in trained animals. This retardation thus seems to outlast even the "physiological" effects. It is therefore unlikely that the deficit is a direct effect of the drug, as is probably the case in the disruption of the learned avoidance response. Instead, some new adaptive response seems to be acquired during the drug phase which persists into Phase II. Especially pertinent in this regard is the high frequency of response failures in the experimental group. This is entirely unlike the frequency for naive animals. What probably happens is that the drugged organism acquires a new mode of adjustment, and this new mode persists after the drug has been metabolized.

That there may be important species differ-

ences has been pointed out by Owens and Walk (5). They reported that reserpine facilitated acquisition of the conditioned avoidance response in white rats. However, dosages, apparatus, shock level, and time between injection and testing differed sufficiently to render comparisons between the results of the two experiments impossible.

SUMMARY

Six *Macacus cynomolgus* monkeys were trained in a shock-avoidance situation while under the influence of reserpine. Three sessions of 60 trials were given a week apart. Subsequent acquisition of the habit was retarded.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533-557.
2. KLINE, N., & STANLEY, A. M. Use of reserpine in a neuropsychiatric hospital. *Ann. N.Y. Acad. Sci.*, 1955, **61**, 85-91.
3. MONROE, R. R., HEATH, R. G., MICKLE, W. A., & MILLER, W. A comparison of cortical and subcortical brain waves in normal, barbiturate, reserpine and chlorpromazine sleep. *Ann. N.Y. Acad. Sci.*, 1955, **61**, 56-71.
4. NOCE, R. H., WILLIAMS, D. B., & RAPAPORT, W.

- Reserpine (Serpasil) in the management of the mentally ill and mentally retarded. *J. Amer. Med. Ass.*, 1954, **156**, 821-824.
5. OWENS, J. W. M., & WALK, R. D. The effect of reserpine on avoidance learning. Abstract of paper read at East. Psychol. Ass., Atlantic City, April, 1956.
6. PLUMMER, A. J., EARL, A. E., SCHNEIDER, J. A., TRAPOLD, J., & BARRETT, W. Pharmacology of *Rauwolfia* alkaloids. *Ann. N.Y. Acad. Sci.*, 1954, **59**, 8-21.
7. RINALDI, F., & HIMWICH, H. W. A comparison of the effects of reserpine and some barbiturates on the electrical activity of cortical and subcortical structures of the brain of rabbits. *Ann. N.Y. Acad. Sci.*, 1955, **61**, 27-35.
8. SCHNEIDER, J. A., & EARL, A. W. Effects of Serpasil on behavior and autonomic regulating mechanisms. *Neurology*, 1954, **4**, 657-667.
9. SMITH, S. K. The use of reserpine in private psychiatric practice. *Ann. N.Y. Acad. Sci.*, 1955, **61**, 206-210.
10. SMITH, R. P., WAGMAN, A., & RIOPELLE, A. J. Effects of reserpine on conditioned avoidance behavior in normal and brain-operated monkeys. *J. Pharm. exp. Therapeut.* 1956, **117**, 136-141.
11. WEISKRANTZ, L., & WILSON, W. A., JR. The effects of reserpine on emotional behavior of normal and brain operated monkeys. *Ann. N.Y. Acad. Sci.*, 1955, **61**, 36-55.

Received May 2, 1956.

TRANSMISSION OF LEARNED BEHAVIOR BETWEEN RATS¹

RUSSELL M. CHURCH

Brown University

THORNDIKE (14) tried to determine whether or not animals are able to learn by observation, i.e., whether or not animals can learn to solve problems by watching a trained animal perform the solution. His results were negative, and his conclusions were that animals learn by trial and error rather than by observation. Thorndike's experiments on observational learning led to a great number of similar experiments. Bird has reviewed most of these studies up to 1940 (2), and Spence has reviewed most of the primate studies up to 1937 (11). A number of the investigators agreed with Thorndike's conclusion that there is no learning by observation in animals, but others claimed to have obtained evidence for this phenomenon. In the latter experiments, the demonstration of the correct response may have increased the activity level of the *S*, or it may have directed his attention to the crucial part of the apparatus (11). Without control of these extraneous variables it is not possible to demonstrate transmission of learned behavior by the observational learning experiment.

Miller and Dollard (7) adopted the position that observation is not sufficient for the transmission of learned behavior between animals but that imitative behavior, like any other behavior, is acquired on the basis of the principles of trial-and-error learning. They demonstrated that hungry rats can learn to follow a leader in a T maze to secure food. These results, also reported by several other experimenters (1, 9), demonstrate that animals can learn to imitate. It was not determined in any of these experiments whether the followers learned the cues to which the leaders had been responding or whether they merely learned to follow the leader. If the followers would respond consistently to the cues to which the

leaders previously had been responding, we may speak of transmission of learned behavior.

The purpose of this experiment is to demonstrate transmission of learned behavior between rats. The design of the experiment treats this as a special case of incidental learning of cues, in which the primary cue is the leader rat. After *S* learns to follow a leader (the primary cue), the leader responds consistently to some new cue (the incidental cue) with *S* still following him. After a number of such trials *S* is placed in the situation *alone*, i.e., without the primary cue of the leader. The consistency to which he then responds to the incidental cue is the index of degree of transmission of learned behavior from leader to follower. On the basis of this analysis, it is clear that this experiment on transmission of learned behavior falls into the class of experiments on incidental learning (3, 4, 6).

METHOD

Apparatus

The apparatus was an elevated T maze (cf. Fig. 1). The stem of the T and each of the arms was 36 in. long. The maze was 32 in. high. The path was wire mesh covering the wood of the maze and was 1.5 in. wide. Two plexiglass starting gates on the stem were located 12 in. and 24 in. from the foot. Two plexiglass goal gates on each arm were located at 12 in. and 24 in. from the choice point. The gates were 12 in. high and 8¾ in. wide. Including the frame into which they were set, the gates extended 4 in. to each side of the path and 4 in. below it. On each of the two arms square metal containers 1.5 in. on a side were sunk 19 in. and 33 in. from the choice point. The lids of these containers were flush with the path. When the lids were raised a thimble of water was exposed from which *Ss* could drink.

Two pairs of 7½-watt red lights were used as discriminative stimuli. One pair was located 5 in. from the choice point on each arm, and the other pair was located on the front starting gate, level with the path.

The room which contained the apparatus was 6 ft. by 10 ft. The entire maze was painted gray, and it was surrounded by a cheesecloth curtain. The room was illuminated by a 10-watt bulb hung 36 in. above the rear starting gate. There was another 10-watt bulb in a desk lamp on *E's* desk which was located behind the foot of the apparatus. During all experimentation a continuous white noise was used to mask extraneous auditory stimulation.

¹ This experiment formed a portion of a dissertation submitted to the faculty of Harvard University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Social Psychology (5). The research was facilitated by the Laboratory of Social Relations, and it was supported by a Ford Foundation Grant-in-Aid to Professor Richard L. Solomon, to whom the author is indebted for aid and advice.

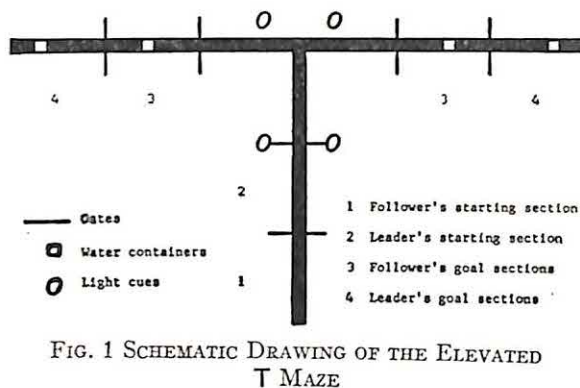


FIG. 1 SCHEMATIC DRAWING OF THE ELEVATED T MAZE

Subjects

Twelve naive male albino rats, about 90 days old, were used in this experiment.

Procedure

Training of the leaders. Three *Ss* were trained to go to the left arm of the T maze and three *Ss* were trained to go to the right arm of the T maze. On each trial *S* was placed in the leader's starting section and the front starting gate was raised. If *S* went to the leader's goal section on the correct arm, he found water in the thimble; if he went to the incorrect arm, he found no water in the thimble. Regardless of the arm he chose, the front goal gate closed behind him (noncorrection method) and he was removed in 30 sec. No *S* was used as a leader who did not consistently reach the correct goal section within 2 sec. after reaching the starting gate.

Training of the followers. On each trial *S* was placed behind the rear starting gate and a leader was placed behind the front starting gate. Both starting gates were raised simultaneously. The leader would run quickly to the choice point, turn to the left (right) arm, and run to the leader's goal section. At that time the front goal gate on both arms would be closed. If *S* went to the same arm as the leader, the lid of the container with water was opened; if he went to the opposite arm, it remained closed. In either case, as soon as *S* reached one of the follower's goal sections, the rear goal gates on both arms would be closed (noncorrection method) and the rat would be returned to his home cage in 15 sec. Each *S* was given 150 trials, 10 trials a day spaced at about 3-min. intervals. Leaders were selected according to a systematic order, such that the sequence of correct trials was LRLRLRLRL, RLRRLRLRL. The *Ss* were housed three in a cage with constant artificial illumination. They were given ad lib food in the home cage, and 30 min. after running each day they were given access to water for 90 min.

Training on the incidental cue. One hundred additional trials were run in which either the two lights on the left of the apparatus or the two lights on the right of the apparatus were lighted. For a given *S* the side on which the light appeared varied in a systematic order. For half the *Ss* leaders were selected who, because of their previous position training, would go to the arm with the light on; for the other half, leaders were selected who would go to the side with the light off. On

each trial when *S* reached either of the follower's goal sections, the lid of the container with water was opened, and *S* was permitted to drink for 15 sec. In all other respects, the procedure for training on the incidental cue was identical to that used in training to follow.

Test for learning of the incidental cue. To test the degree of learning of the incidental cue, the procedure for the training on the incidental cue was continued *except that there was no leader*. Thus, on each trial the lights were on the left (or right) and the follower was placed in the rear starting section. The gates were raised, and regardless of the arm to which *S* went, he was given water for 15 sec. in the follower's goal section. Ten such test trials were run on each of two days. The number of times *S* chose the arm marked by the cue which had been positive for the now-absent leader was recorded as the measure of the amount of transmission of learned behavior.

RESULTS

On the initial 10 trials there was no significant tendency for *Ss* to go to the same arm of the T maze as the leader or to go to the opposite arm. The *Ss* gradually learned to follow the leader. On the last 20 trials the six *Ss* followed the leader on all but one trial. (cf. Fig. 2).

On the initial trials after the introduction of the incidental cue there was a slight decrease in the percentage of following responses (external inhibition?). Despite the fact that *Ss* could obtain water on either arm of the maze during this period, they followed the leader on 95 per cent of the trials. A few *Ss* developed side preferences during this treatment, thus slightly decreasing the total percentage of following responses.

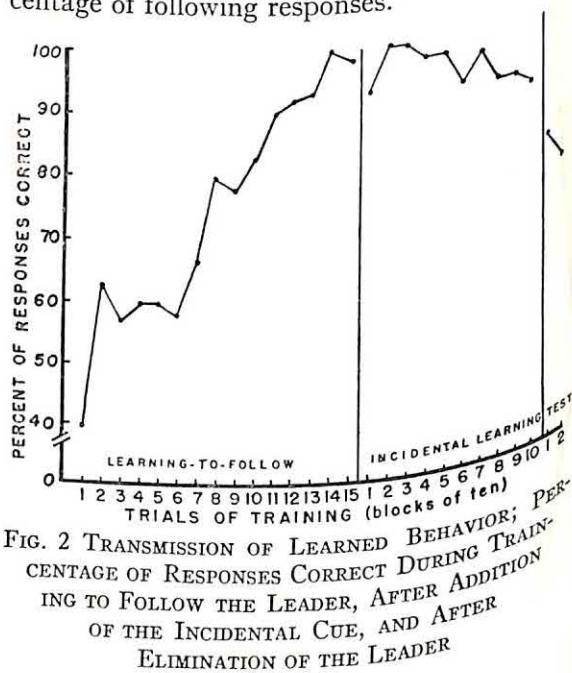


FIG. 2 TRANSMISSION OF LEARNED BEHAVIOR; PERCENTAGE OF RESPONSES CORRECT DURING TRAINING TO FOLLOW THE LEADER, AFTER ADDITION OF THE INCIDENTAL CUE, AND AFTER ELIMINATION OF THE LEADER

After the leaders had been omitted, Ss went to the arm marked by the cue to which the leader had been going on 77 per cent of the 20 test trials ($t = 4.4$; $p < .01$). These test trials demonstrate that Ss learned the cues to which the leaders were responding.

DISCUSSION

Miller and Dollard (7) have demonstrated that imitative behavior may be acquired on the basis of trial-and-error learning. The present experiment further confirms this hypothesis, and it also presents evidence that such imitative behavior can become *independent* of the behavior of the leader. In this experiment rats acquired a discrimination habit in a T maze after following leaders who had acquired that habit. Such transmission of learned behavior is to be expected on the basis of a continuity theory of discrimination learning (12). According to the continuity theory, a set to respond to one aspect of the stimulus situation does not interfere with the learning of other effective cues preceding the reward. Thus, in the present case, a set to follow the leader should not interfere with the learning of the incidental cue preceding the reward. Although there is some evidence to the contrary (6) most evidence supports the hypothesis that incidental cues may be learned (3, 4, 10).

The present laboratory demonstration does not duplicate the naturalistic situation, but it indicates a mechanism through which social transmission of behavior could readily occur. The positive findings encourage speculation about the possibility of a nonpurposive transmission of culture among animals. According to the hypothesis, a rat who has learned to follow a leader will learn to respond to the cues to which the leader has responded. If another rat learned to follow him, he, too should learn to respond independently to these cues. In this case, behavior learned by an original leader would have been transmitted through two "generations." Although the necessary condition of following behavior has been demonstrated in certain animal species (8, 13), experiments in field situations have not yet been performed to determine whether transmission of learned behavior occurs in nature.

SUMMARY

Six rats were trained to follow leader rats in an elevated T maze to secure a reward of water. An incidental cue of two lights was then introduced such that the leaders responded consistently with respect to it. After 100 trials of following leaders who were responding to the incidental cue, Ss were given 20 trials alone. On 77 per cent of these test trials Ss went to the arm marked by the cue to which the leader had been going. This finding was interpreted as a demonstration of transmission of learned behavior between animals, and it was explained on the basis of a continuity theory of discrimination learning.

REFERENCES

1. BAYROFF, A. G., & LARD, K. E. Experimental social behavior of animals III. Imitational learning of white rats. *J. comp. physiol. Psychol.*, 1944, **37**, 165-171.
2. BIRD, C. *Social psychology*. New York: Appleton-Century, 1940.
3. BLUM, R. A., & BLUM, J. S. Factual issues in the 'continuity' controversy. *Psychol. Rev.*, 1949, **56**, 33-40.
4. BRUNER, J. S., MATTER, J., & PAPANAK, M. L. Breadth of learning as a function of drive level and mechanization, *Psychol. Rev.*, 1955, **42**, 1-10.
5. CHURCH, R. M. Factors affecting learning by imitation in the rat. Unpublished doctor's dissertation, Harvard Univer., 1956.
6. LASHLEY, K. S. An examination of the 'continuity theory' as applied to discrimination learning. *J. Gen. Psychol.*, 1942, **26**, 241-265.
7. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
8. SCOTT, J. P. Social behavior, organization, and leadership in a small flock of domestic sheep. *Comp. psychol. Monogr.*, 1945, **18**, No. 4.
9. SOLOMON, R. L., & COLES, M. R. A case of failure of generalization of imitation across drives and across situations. *J. abnorm. soc. Psychol.*, 1954, **49**, 7-13.
10. SPENCE, K. W. An experimental test of the continuity and noncontinuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253-266.
11. SPENCE, K. W. Experimental studies of learning and the higher mental processes in infra-human primates. *Psychol. Bull.*, 1937, **34**, 806-850.
12. SPENCE, K. W. The nature of discrimination learning in animals. *Psych. Rev.*, 1936, **43**, 427-449.
13. STEWART, J., & SCOTT, J. P. Lack of correlation between leadership and dominance relationships of goats. *J. comp. physiol. Psychol.*, 1947, **40**, 255-264.
14. THORNDIKE, E. L. Animal intelligence. *Psychol. Rev. Monogr. Suppl.*, 1898, **2**, No. 4.

Received April 23, 1956.

EFFECT OF ANXIETY AND TWO KINDS OF MOTIVATING INSTRUCTIONS ON VERBAL LEARNING¹

IRWIN G. SARASON

University of Washington

THIS experiment investigated the relationship between anxiety and differential motivating instructions in a verbal learning situation. Anxiety was here defined by scores obtained by subjects (Ss) on the Taylor Manifest Anxiety Scale (MAS) and two sets of motivating instructions were used. By means of one set, subject-oriented instructions, Ss were told they were about to take an intelligence test; by means of the other set, experimenter-oriented instructions, Ss were urged to do well as a favor to the experimenter. In addition to studying groups in which these kinds of instructions were used, control groups, receiving no motivating instructions, were also studied. Two questions stimulated the present research: (a) How does anxiety affect learning when the variable of verbally-induced motivation is manipulated? (b) If anxiety does interact with motivational instructions, does the interaction depend upon the content of such instructions?

Up until recently, most textbook writers and researchers in the area of motivation have suggested that the effect of motivating communications to Ss is to impel them to greater effort and hence to superior performance. Further research, however, has indicated that the effects of instructions are considerably more complicated than this broad generalization would suggest (6, 7, 9, 10, 11). It was found, for example, that motivating instructions, rather than facilitating performance, resulted in large intersubject variability (6). Sarason (9), inferring that this increased variability among Ss was in part attributable to individual personality differences, performed an experiment in which, in addition to instructions, anxiety was also manipulated. It was found that anxiety, indeed, did interact with instructions, but that with this variable taken into account no significant concomitant increase in intersubject variability was obtained. Sarason found that highly motivating instructions were facilitative only for middle-

anxious (MA) and low-anxious (LA) groups (anxiety defined by MAS scores). On the other hand, the high-anxious (HA) groups given the instructions performed at a significantly lower level than did a HA control group.

Since this finding has important implications for current theories of anxiety, it was felt that future experiments which attempted to replicate these latter findings under varying conditions would be of considerable relevance, particularly with regard to generalizations which might be made concerning the effect of anxiety on behavior. In the present study, aside from the addition of another type of instructional condition, there was only one change in the experimental conditions from those used in the earlier research. Whereas in the earlier experiment, nonsense syllables of 73 per cent association value were used, this experiment used syllables of 53 per cent association value. It was tentatively assumed that decreasing the association value of the syllables would make the list to be learned more difficult. Replicating the earlier experiment with more difficult material seemed of interest in light of the relationship between anxiety and difficulty of task hypothesized by Hullian drive-theorists (3, 13). This theory predicts that HA Ss are superior to LA Ss on relatively easy tasks, but the reverse holds on more difficult tasks. Difficulty or complexity was defined in terms of the number of competing responses in S's response repertory for the purposes of this experiment. While the present experiment did not manipulate the variables of task difficulty and complexity, it does provide a basis for evaluating the generality of previous findings.

In addition to evaluating the interactive effects of anxiety and motivating instructions, this experiment also studied the efficacy of two kinds of motivating instructions. As is already known, certain instructions may be facilitative for some groups of Ss but not for others. The kinds of instructions which have had a deleterious effect on HA Ss were those in which S was more or less placed "on the

¹ This research was performed while the author was a predoctoral USPHS fellow at Indiana University.

spot" to attain a high level of achievement. Would HA Ss be less detrimentally affected if they were urged to do well, not for the purpose of preserving their own self-esteem, but rather as a means of aiding a graduate student? It was predicted that the subject-oriented instructions would result in greater decrements in performance for HA Ss than would experimenter-oriented instructions.

METHOD

Subjects. The 75 Ss in this experiment were drawn from the upper, middle, and lower 20 per cent of the MAS distribution for 530 introductory psychology students at Indiana University. The median for the distribution was 16. The 21 HA Ss had MAS scores above 24, while the 24 LA Ss had scores below 9. The MA Ss had scores between 14 and 18. All Ss were naive with regard to experiments on serial learning. No Ss with *T* scores above 70 on the MMPI *K* scale were used.

Procedure. A Hull-type memory drum was used to expose lists of 17 nonsense syllables with a 2-second exposure per syllable and a 6-second intertrial interval. The anticipation method was employed.

The syllables used were selected from Glaze's lists (4). The 17 nonsense syllables of List A, the preliminary list in this experiment, included 6 syllables randomly selected from Glaze's list of 27 per cent association value, 6 syllables randomly selected from Glaze's 80 per cent list, and 5 syllables from the list of 53 per cent association value. The experimental list, List B, consisted of 17 syllables randomly selected from Glaze's list of 53 per cent association value, with the restriction that no syllable included in List A could be used in List B.

All Ss were initially given 15 trials on List A. The preliminary instructions used were the same as those described by Sarason (9). After performance on List A, Ss in each anxiety group were randomly assigned to one of three experimental groups. These groups differed with respect to the instructions they received prior to beginning List B. One group received subject-oriented instructions in which *S* was told that the next list of nonsense syllables (List B) was a short form intelligence test and his score would be compared with those of other Ss. These instructions were called subject-oriented because pressure was placed on *S* to maintain his self-esteem by performing at a high level. The subject-oriented instructions used were the same as those described by Sarason (9).

The second kind of motivating instructions studied was called experimenter-oriented. Here emphasis was placed on the "tight spot" in which *E* found himself. These instructions were as follows:

I know it isn't easy learning these syllables. This experiment is part of my thesis and getting my degree depends on getting good results in this experiment. I would appreciate it a lot if you would do your best to memorize these syllables as soon as you can.

The control groups received neutral instructions which stated that the next list was a practice list and

that *S*'s performance was not the main concern; rather, *E*'s interest was in discovering certain characteristics of nonsense syllables. These neutral instructions were the same as those used by Sarason (9).

Following the differential instructions, Ss received 20 trials on List B. At the end of Trial 20, Ss were asked to return 24 hours later. At that time Ss were given 10 more trials on List B.

During the experiment *E* was unaware of the anxiety scores of the Ss, and he did not know which instructions *S* was to receive until the 15 trials of List A had been completed.

RESULTS

In order to assess the effects of anxiety prior to administration of differential motivating instructions, a 3×3 analysis of variance was performed on the total number of correct anticipations on List A. None of the *F*s obtained in this analysis approached significance. It was concluded that anxiety scores were not shown to be related to performance prior to administration of motivating instructions. It seemed reasonable, therefore, to interpret significant findings on List B in terms of the experimental treatment.

The results on the experimental list, List B, for learning (Day 1) and relearning (Day 2) were analyzed separately. Although no group differences were significant on List A, analyses of covariance were used on the experimental data in order to control for individual differences in learning ability and thus increase the precision of the experiment. The control variable used in each covariance analysis was the total number of correct anticipations on List A. The criterion measure used on Day 1 was the total number of correct anticipations for the first 20 trials of List B. On the Day 2 analysis, the criterion measure was the total number of correct anticipations for Trials 21-30.

The analysis of covariance for Day 1 yielded significant *F*s for the anxiety variable, the instructional variable, and the interaction of these two variables. Table 1 presents the means, standard deviations, and adjusted means for performance on List B on Day 1. Table 2 summarizes the results of the analysis of covariance for this data. Examination of Table 1 reveals that the HA and LA groups performed at a significantly higher level than did the MA group. This relatively poor performance for the MA Ss was due primarily to the low level of the MA control group's performance. This group was found to have

TABLE 1

MEANS, *SD*s, AND ADJUSTED MEANS FOR NUMBER OF CORRECT ANTICIPATIONS ON TRIALS 1-20 OF LIST B

Anxiety Level	Instructions								
	Neutral			Subject-Oriented			Experimenter-Oriented		
	Mean	<i>SD</i>	Adj. Mean	Mean	<i>SD</i>	Adj. Mean	Mean	<i>SD</i>	Adj. Mean
HA	151.1	25.6	140.1	115.6	41.1	114.8	129.9	38.7	126.6
MA	87.9	24.5	90.7	115.7	45.9	123.2	113.6	38.2	117.9
LA	110.9	39.1	105.2	151.1	15.8	154.1	130.0	63.1	127.8

TABLE 2

ANALYSIS OF COVARIANCE OF NUMBER OF CORRECT ANTICIPATIONS ON TRIALS 1-20 OF LIST B

Source of Variation	<i>df</i>	Mean Square	<i>F</i>
Anxiety	2	2987.65	3.28*
Instructions	2	2869.09	3.15*
A × I	4	2740.62	3.07**
Within Groups	65	911.09	
Total	73		

* Significant at .05 level.

** Significant at .025 level.

given significantly fewer correct anticipations on the first 20 trials of List B than did the MA subject-oriented and experimenter-oriented groups ($p < .02$).

The significant main effect for instructions revealed an over-all superiority of the subject- and experimenter-oriented groups to the neutral control groups. However, within the control or neutral condition alone, the HA Ss attained the highest performance level. This HA group was found to be significantly superior to the LA as well as the MA groups similarly treated ($p < .001$). There were no significant differences among the three anxiety groups under the experimenter-oriented condition. The anxiety × instructions interaction which was obtained was due to the superiority of the control HA group to the two HA groups given the motivating instructions, and also to the superiority of the two kinds of motivating instructions relative to the neutral condition for the MA and LA Ss. Although the subject-oriented instructions tended to result in better performance for both the MA and LA Ss than did comparable experimenter-oriented instructions, this superiority was found to be statistically significant only for LA Ss. The subject-oriented LA Ss attained the highest level of performance of all Ss in the experiment on Day 1.

Although Bartlett's test did not result in a significant chi square, it was noted that there was considerable more intersubject variability than in previous experiments using similar experimental situations. This result raises the question of whether or not control of only one individual difference variable, anxiety, is sufficient to reduce markedly the intragroup variability so often obtained when motivating instructions are experimentally varied.

Fig. 1 shows the learning curves of the three anxiety groups under the three instructional conditions for blocks of 5 trials for Trials 1-20 of List B. Examination of the curves in Fig. 1 clearly demonstrates the marked superiority of the HA group to the other two anxiety groups under the neutral condition. Under the subject-oriented condition, the LA group improves much faster than do the HA and MA groups. Interestingly, the differences among the anxiety groups under the experimenter-oriented condition are much smaller than under the other instructional conditions.

The significant over-all differences among the three anxiety groups obtained on Day 1

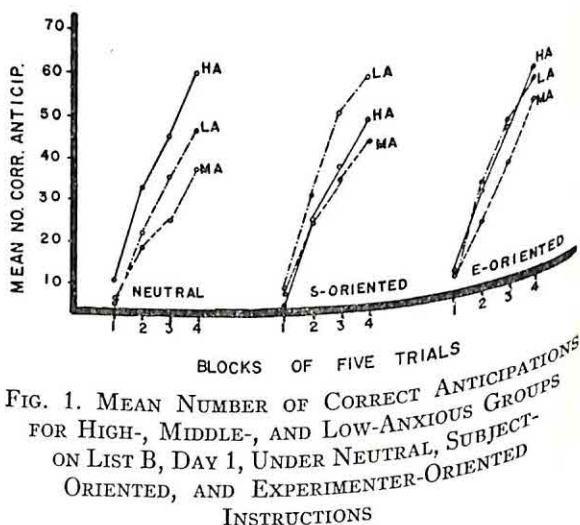


FIG. 1. MEAN NUMBER OF CORRECT ANTICIPATIONS FOR HIGH-, MIDDLE-, AND LOW-ANXIOUS GROUPS ON LIST B, DAY 1, UNDER NEUTRAL, SUBJECT-ORIENTED, AND EXPERIMENTER-ORIENTED INSTRUCTIONS

were also found to hold on Day 2. Both the HA and LA Ss were found to give a significantly greater number of correct anticipations than did the MA group ($p < .05$), the difference between the HA and MA groups ($p < .001$) being considerably greater than the difference between the LA and MA groups ($p < .05$).

There were two other noteworthy findings. The first involved the HA control group. This group had a significantly greater number of correct anticipations on Day 2 than all other groups in the experiment. The other striking finding was the apparent dissipation of the facilitative effect of subject-oriented instructions observed on Day 1. It is interesting that the means for all three LA groups were identical on Day 2. Among the MA Ss, the neutral group continued to perform more poorly than did the groups given motivating instructions. In this connection, it should be observed that the anxiety $\times$ instructions interaction, while not significant, was in the same direction as found in previous experiments (9, 11).

DISCUSSION

The finding on the preliminary list of no difference between HA, MA, and LA groups is in line with previous evidence of no effect for anxiety prior to administration of experimental instructions (9, 10). This result, together with data of the postinstruction performance of HA Ss given the neutral and motivating instructions, lends support to the hypothesis that HA Ss differ from Ss in the rest of the anxiety-score distribution in the extent to which they are deleteriously affected by stressful communications. In a relatively complex situation such as that confronting Ss in the present experiment, their performance markedly declines only when given such communications. This result would appear to be more consistent with a habit (9) than with a general drive (3) interpretation of anxiety.

One aim of the present experiment was to compare the effect on performance of motivating instructions which differ in that one set informs *S* that his intellectual capacity is being tested, whereas the other merely asks *S* to help a fellow student. It had been anticipated that HA Ss would not show the detrimental effect of motivating instruction under the experimenter-oriented condition.

The results both on Day 1 and Day 2, while not always attaining statistical significance, tended to support this expectation. It was shown that both under the experimenter-oriented and under the subject-oriented conditions the performance of the HA Ss were deleteriously affected. In this connection it is interesting to note that even success communications to HA Ss can have a detrimental effect (8). However, the detrimental effect of the experimenter-oriented instructions on HA Ss tended to be less than the effect obtained with subject-oriented instructions. This result also would appear to support the hypothesis that experimenter-oriented instructions are less anxiety provoking than are subject-oriented instructions.

For all MA and LA Ss combined, the subject-oriented instructions led to better performance than did the experimenter-oriented instructions. This result was highly significant for the LA group and the findings for the MA group, although not approaching significance, were in the same direction. If, therefore, one is interested in the hypothesis that motivating instructions are facilitative for LA and MA Ss and detrimental for Ss who obtain HA scores, it would appear that the experimenter-oriented instructions used in this study are probably more motivating than neutral instructions and less motivating than subject-oriented instructions. The development of such gradations in differential instruction would seem to be of considerable usefulness in studying the effects of experimentally manipulated motivation on behavior.

This experiment was designed to permit comparison with an earlier investigation by Sarason (9). As found in that study, HA Ss in the present experiment were detrimentally affected by the motivating instructions, whereas LA and MA Ss were facilitated in their learning by the motivational instructions. This interaction was highly significant on Day 1, and while not significant on Day 2, it was in the expected direction. How might this difference in reaction to experimentally induced anxiety be explained? A probably fruitful area for investigation in seeking an answer to this question is that of *S*'s interpretation of presumably stressful situations. To the MA and LA *S* the motivational instructions may merely have functioned as

an incentive and thus resulted in the facilitative effect which is so often attributed to motivational instructions. However, to the HA Ss, who may at the outset of the experiment have strong doubts about his adequacy in novel situations, such instructions may have aroused strong fears of failure. Self-verbalizations about his inadequate performance, as well as physiological concomitants of anxiety may be important variables detrimental to the performance of HA individuals. In any event, the difference between HA, MA, and LA Ss does not appear to be a general one which affects all behavior indiscriminately. Rather it appears that HA Ss differ from all other Ss in anxiety-score distribution in that their performance is detrimentally affected by stressful, evaluative communications. Carrying this finding one step further involves systematic attempts to delineate variables which are related to the performance of HA individuals. Available experimental data suggest that such factors as socioeconomic status (12) and sex of S (2) have a definite relationship to the performance of Ss differing in anxiety scores.

If HA Ss differ from MA and LA Ss only in their reaction to experimentally induced anxiety, one would expect no effect for anxiety prior to administration of experimental instructions, as well as no difference under a control or neutral condition. Whereas the former expectation was confirmed, the latter was not. Both on the immediate and on the delayed (24 hours) test of retention, the HA Ss not given motivating instructions were significantly superior to MA and LA Ss similarly treated. The fact that this superiority of the control group HA Ss occurs on List B and not on List A may suggest that HA Ss under nonstressful conditions, may profit more than do MA and LA Ss from familiarity and practice with the task to be performed. In similar experiments, Sarason (9, 10) has reported essentially the same finding, i.e., under neutral conditions HA Ss make a greater improvement in performance than either MA or LA Ss. Ausubel *et al.* (1) have recently reported a finding like this for HA and LA groups where anxiety was defined by certain Rorschach test indices. This type of result leads one to speculate on the role of hypothetical task-irrelevant habit tendencies, e. g.,

saying to oneself, "I can't do this test," which interfere with the performance of HA individuals. Possibly increased practice and familiarization in the task lead to a decrement of these kinds of responses under neutral conditions.

In this connection it is significant that both Sarason and Mandler (12) and Gordon (5) have found that HA students (defined by score on the Test Anxiety Questionnaire) obtain significantly poorer scores on college entrance examinations than do students who have LA scores, but that just the opposite is the case for course grade-point averages. This result raises the possibility that HA individuals when given sufficient time for familiarization with the task required of them (e.g., learning the course material) not only do as well as other students but actually may surpass them. This consideration leads to an important question: Is the relatively greater improvement in performance of HA Ss a function merely of habituation to the task at hand, or persistence, or of extinction of the interfering responses that have been hypothesized?

The present result of significant over-all differences due to anxiety alone (HA and LA groups superior to the MA group) is somewhat deviant from previous findings and expectations. Certainly, the relatively poor performance of the MA Ss and the relatively good performance of the HA Ss (especially on Day 2) are somewhat surprising. Neither the recently propounded drive or habit theories would have predicted this result. A drive interpretation (3, 13) of anxiety would probably have expected differences among the three anxiety groups, but with LA Ss superior to MA Ss, and the latter superior to Ss who are HA. This unsupported hypothesis presumes differences in drive level among the three groups and the relative difficulty of the task used in this experiment. Although the variable of task difficulty was not manipulated in this experiment, it seems reasonable to assume that the list of 17 nonsense syllables provided S with a large number of competing responses. A habit theory (8, 9), emphasizing differences between HA Ss and all other Ss in the anxiety-score distribution in terms of competing interfering responses, also cannot adequately explain the results with regard to the three anxiety groups since this theory as

thus far formulated merely states that under neutral conditions, HA Ss should not perform differently than other Ss, and that differences should appear only under stressful conditions. While in general, the results of this experiment support this theory, the theory provides no basis for expecting a superiority of HA Ss.

The present results, then, generally support the hypothesis that differential motivating instructions can be meaningfully and reproducibly manipulated. The present experiment in several important aspects has reproduced earlier findings obtained under similar and controlled conditions. It would appear important at this point to manipulate such variables as the sex and personality characteristics of the experimenter (administrator of instructions) in order to gain a still firmer understanding of the effect of motivating instructions on performance.

SUMMARY

The present experiment studied the effects of anxiety (Taylor Manifest Anxiety Scale scores) and differential motivating instructions on serial learning and retention of nonsense syllables by college students. One kind of motivating instructions, called subject-oriented, emphasized the need for *S* to perform well in order to maintain his self-esteem. Another kind of motivating instructions, experimenter-oriented, involved enlisting *S* to help *E* by performing well. Neutral control groups were also studied.

Anxiety, motivating instructions, and the interaction of these two variables, were found to have significant effects on learning. Only anxiety had a significant effect when retention

was measured 24 hours later. The results were discussed in terms of a habit interpretation of anxiety.

REFERENCES

1. AUSUBEL, D. P., SCHIFF, H. M., & GOLDMAN, M. Qualitative characteristics in the learning process associated with anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 537-547.
2. AXELROD, H. S., COWEN, E. L., & HEILIZER, F. The correlates of manifest anxiety in stylus maze learning. *J. exp. Psychol.*, 1956, **51**, 131-138.
3. FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311-327.
4. GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255-267.
5. GORDON, E. M. Self-esteem, need for achievement, test anxiety, and performance. Unpublished doctor's dissertation, Yale Univer., 1954.
6. LAZARUS, R. S., & ERIKSEN, C. W. Effects of failure stress upon skilled performance. *J. exp. Psychol.*, 1952, **65**, 59-66.
7. LUCAS, J. The interactive effects of anxiety, failure, and intra-serial duplication. *Amer. J. Psychol.*, 1952, **65**, 59-66.
8. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166-173.
9. SARASON, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253-260.
10. SARASON, I. G. The effect of anxiety and two kinds of failure reports on serial learning. *J. Pers.*, in press.
11. SARASON, I. G. The effect of association value and differential motivating instructions on serial learning. Unpublished manuscript.
12. SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810-817.
13. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety to level of performance in serial learning. *J. exp. Psychol.* 1952, **44**, 61-64.

Received May 1, 1956

SOCIAL PRESSURE IN PUBLIC AND PRIVATE SITUATIONS

MICHAEL ARGYLE

University of Oxford, England

NORM FORMATION is one of the simplest and most widely studied phenomena in connection with small social groups. The experiments of Festinger (3) and his group have established some of the conditions under which it takes place. It has been argued elsewhere (2) that the most useful way of explaining group phenomena is by means of postulates concerning *individual* behavior, and a deductive theory of this type has been put forward (1) to account for norm formation. It sets out from Schachter's finding (7) that deviates are rejected, postulating also that there is a need for acceptance by individuals and groups, and that people learn from experience to conform in order to be accepted. A number of predictions follow from the theory, one of which is tested here. This is that less social influence occurs when the behavior in question is not observable by those exerting the pressure—i.e., when it takes place in private, than when the behavior is public. In the experiment reported here Ss were exposed to an identical amount of social pressure, but with different degrees of exposure to those exerting the influence. It also follows from the theory that no influence at all occurs in the private condition. Obviously some extension is required to deal with "internalization" processes, but this refinement is disregarded here.

Another deduction from the theory is that more influence occurs when the S is highly motivated to be accepted by the influencing individual or group. The fact that people conform to the norms of reference groups (6)—in the sense of "groups to which they wish to belong"—may be regarded as confirmation of this. In this experiment an attempt was made to manipulate the state of Ss' "need for acceptance," in order to provide further confirmation.

There are thus two hypotheses to be tested here: (a) There will be more social influence in the Public than in the Private condition. (b) There will be more social influence when the need for acceptance is aroused.

METHOD

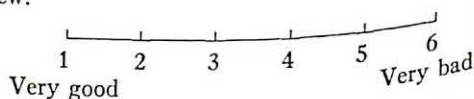
Subjects

The experiment was first carried out with 27 boys from the 5th and 6th forms of the Cambridgeshire High School. It was later repeated with 25 male students from Ruskin College, Oxford.

General Procedure

Two Ss, who are initially strangers, are introduced and seated at opposite sides of a table three feet apart, with a screen between them so that they cannot see one another. They are then handed the following instructions.

This is an experiment to study social communication under somewhat unusual conditions. We want you to discuss the picture which you will be shown, not by talking, but by passing messages on the slips of paper provided. First, will you make up your mind what your initial judgment of the picture is, using the following scale, and mark on the scale at the point which best represents your view.



For your first message will you send a message simply stating the number on the scale corresponding to your judgment of the picture. When you have both done that will you carry on a discussion of the picture by passing further message slips, and try to come to some agreement in ten minutes. There is no restriction on the kind of messages sent—you can send absolutely anything you like.

The picture in question was Chagall's "Poète Allongé," chosen for its unusual and ambiguous character. When four messages had been exchanged the experimenter stopped the discussion and gave out a questionnaire (described below).

Creating a Standard Amount of Social Influence

In the second phase of the experiment, when students were used, one of the "Ss" was actually a confederate who copied out and sent a series of standard messages to the S. The confederate waited for S to send the first message—containing the scale number of his judgment of the picture—and sent "1" or "6," whichever was the extreme opposite of S's judgment. As a 6-point scale was used, with no center point, all Ss had to decide one way or the other. Virtually the same series of messages was used either to attack or to defend the picture, e.g.
For: "What you say is trivial, for the picture is so meaningful as a whole."

Against: "What you say is so trivial, for the picture is so meaningless as a whole."

In the first phase of the experiment, when school-boys were used, real Ss were in pairs, but the messages were intercepted by the experimenter and standard ones substituted. The confederate arrangement worked more smoothly, and required less sleight of hand.

Manipulation of Experimental Variables

The *public/private dimension* (*Pu/Pr*) was manipulated by arranging two contrasted ways of expressing the final judgment. In the *Pu* condition, Ss got up from behind the screen and told each other face to face. In the *Pr* condition they entered their final judgment on a questionnaire and placed this in a sealed envelope; they were assured that the contents would be treated statistically and would not be connected with them personally.

The *need for acceptance* was varied by giving an "accepting" (*A*) or "rejecting" (*R*) tone to the message sent to the subject, e.g., *A*. "I respect your opinion, but the picture is so meaningless as a whole." *R*. "What you say is trivial, for the picture is so meaningless as a whole." It was assumed that the need for acceptance would behave like a biological drive so that deprivation should lead to increased drive strength. Thus rejection here should increase the level of the need for acceptance and lead to greater social influence.

The *status* of the confederate in relation to the *S* was manipulated in two alternative ways in the two phases of the experiment. Since this manipulation produced no effect on the dependent variables it will be omitted from discussion.

Measurement of Dependent Variables

The principal dependent variable was the occurrence and extent of change of judgment. This was measured by a comparison of initial and final judgments on the same 6-point scale. The Ss were also asked "What would you say was the attitude of the other person to yourself? (it is realized that this is difficult to answer, but make a guess at what seems most appropriate)." A 7-point scale from "dislikes me very much" to "likes me very much" was provided. This measure was introduced to check the effectiveness of the *A/R* manipulation. Since the scales must be assumed to be ordinal, the results have all been computed as frequencies, and tests for significance have been made with χ^2 .

RESULTS

The results of both phases of the experiment have been combined here: both show the same trends.

Influence in Public and Private Conditions

As shown in Table 1, 38.5 per cent of the Ss were influenced by the persuasion in the expected direction, and 7.7 per cent reacted in the opposite direction. Comparing those who

changed in the expected way with all the others, there was significantly more influence in *Pu* than in *Pr*, confirming the first hypothesis. (Chi square = 4.00 [using Yates's correction], $p < .05$). It will also be noted that in *Pu* three Ss reacted in the opposite direction, as opposed to one in *Pr*. This finding is in the opposite direction to our general prediction, though the numbers are far too small to be significant, and may easily be due to chance effects.

Influence in Accepting and Rejecting Conditions

The second hypothesis—that there would be more social influence in *R* than in *A*—was not confirmed, as can be seen by inspection from Table 1. Furthermore, there was no relation between influence and *perceived* rejection, as Table 2 shows.

The measure of *perceived* rejection was intended to provide a check on the manipulation of the *A/R* variable—assuming that perceptual defense was not occurring. As is shown in Table 3, there was more *perceived* rejection in the *R* condition (chi square = 6.43, $p < .05$ with 2 *df*). However while this relationship also holds within the *Pr* condition examined separately ($p < .04$), there is only a tendency in the expected direction for the *Pu* condition.

TABLE 1
PERCENTAGES OF SUBJECTS CHANGING THEIR OPINION
IN EACH CONDITION
(Numbers given in brackets)

Condition	Direction of Change		
	Toward Partner	No Change	Away from Partner
Pu A	57.1 (8)	35.7 (5)	7.1 (1)
Pu R	50 (6)	33.3 (4)	16.7 (2)
Pr A	25 (3)	66.7 (8)	8.3 (1)
Pr R	21.4 (3)	78.6 (11)	0 (0)
Total	38.5 (20)	53.8 (28)	7.7 (4)

TABLE 2
INFLUENCE AND PERCEIVED REJECTION

Frequency	Perceived rejection (1-3)	Midpoint (4)	Perceived acceptance (5-7)
Numbers influenced	6	8	6
All others	13	11	8

TABLE 3
PERCENTAGES OF SUBJECTS WHO PERCEIVED THEMSELVES TO BE REJECTED IN EACH CONDITION
(Scale points 1-3 of the 7-point scale are taken to indicate perceived rejection.)

Condition	Condition	
	Accepting	Rejecting
Public	21	33
Private	17	64
Total	19	50

General Observations

Although no formal interview was undertaken at the end of the experiment, the spontaneous comments of the Ss were of interest. A minority realized that they were expected to change their opinion and so actively resisted this; others thought that it was a telepathy experiment. Many Ss found the "perceived rejection" question difficult, and some of those in the *R* condition felt that it was their opinion rather than themselves which was being rejected, and so did not report rejection.

DISCUSSION

The greater amount of influence in *Pu* as opposed to *Pr* is as predicted, and confirms the theory that norm formation is connected with the need for acceptance. No greater influence was found in *R* than in *A*, contrary to our prediction. One reason for this may be that the *A/R* variable was not manipulated effectively. The measure of *perceived* rejection indicated that the *A/R* variable was successfully changed in *Pr* but not in *Pu*. Examination of the influence results for *Pr* alone fails to show any significant difference between *A* and *R* (see Table 1). Therefore it may be presumed that ineffective manipulation is not the explanation.

A second explanation of our results may be that *A* and *R* do not create changes in the need for acceptance. McClelland (5, p. 100) found that the achievement motive was not influenced by the experience of success and failure unless Ss had previously been given suitable ego-involving instructions to sensitize them to the need in question. It may be that the need for acceptance behaves in this way rather than in accordance with the simple biological deprivation model postulated here.

Thirdly, it is possible that there is no drive

to be accepted by people in general, but only by particular people and groups, who resemble those connected with gratification in the past. That there is a difference here is demonstrated by experiments on reference groups. To test this possibility the experiment could be repeated, using confederates who are friends or high-status colleagues of the Ss.

Lastly, it should be noted that there are psychological grounds for expecting that rejection would lead to *less* influence, not more. Gardner (4) found that if *A* likes *B*, *A* is more influenced by *B*. If it is the case that when *B* rejects *A*, *A* will reject *B* in turn, then it follows that rejection leads to less influence. The present negative result may be due to a confusion of this process with the predicted one.

Perceived rejection was significantly greater in *R* than in *A*—for the *Pr* condition only, a tendency in the same direction being found for the *Pu* condition. One explanation of the ineffectiveness of the *A/R* variable in *Pu* might be as follows: some of the confederates felt unhappy about sending the *R* messages, and they may have been unintentionally friendly in order to compensate during the public announcement. The brief face-to-face encounter in *Pu* may have been enough largely to destroy the effect of the series of rejecting messages.

SUMMARY

Fifty-two groups of two were set up experimentally, containing one genuine *S* and one confederate. The *S* received standard sets of written messages, in what was ostensibly a discussion of the merits of a picture.

The Ss manifested more social influence when their final opinion was stated in "public" to the confederate, than when stated in "private."

It was predicted that more influence would occur when the messages had a "rejecting" tone than when they were "accepting." This did not occur; various explanations are advanced.

Perceived rejection was greater when rejecting messages were received. This effect was most marked in the "private" condition.

REFERENCES

1. ARGYLE, M. Deductive theories in sociology. *Sociol. Rev.*, 1956, 3, 219-234.

2. ARGYLE, M. *The scientific study of social behavior*. London: Methuen, 1957.
3. FESTINGER, L., ET AL. *Theory and experiment in social communication*. Ann Arbor: Institute for Social Research, Univer. of Michigan, 1950.
4. GARDNER, G. Functional leadership in small groups. Unpublished doctor's dissertation, Univer. of London, 1956.
5. MCCLELLAND, D., et al. *The achievement motive*. New York: Appleton-Century-Crofts, 1955.
6. NEWCOMB, T. M. *Personality and social change*. New York: Dryden, 1943.
7. SCHACHTER, S. Deviation, rejection, and communication. *J. abnormal. soc. Psychol.*, 1951, **46**, 190-207.

Received April 13, 1956.

A FACTOR-ANALYTIC STUDY OF RELIGIOUS ATTITUDES¹

WILLIAM E. BROEN, JR.

University of Minnesota

THE existence of rather widespread beliefs concerning the importance of religious attitudes as psychological phenomena is manifested by the large number of studies in this area. Although there have been many investigations of the correlates of religious attitudes it is likely that these studies have not been maximally fruitful because of lack of knowledge about how religious attitudes should be analyzed. Are religious attitudes unidimensional or multidimensional? If multidimensional, what are these dimensions and how can they be measured?

Several factor-analytic studies, which were designed primarily to explore the area of broad general attitudes rather than being specific to the area of religious attitudes, have touched on these questions (1, 2, 3, 5, 6). In general, these studies tended to characterize religious attitudes as unidimensional with that dimension being described either as a "religiosity" dimension or as part of a broader "radicalism-conservatism" factor. Because all of these studies were done on rather restricted populations (college students or graduates), and most used only a few scales relevant to the specific area of religious attitudes, it was felt that additional study was needed prior to the giving of definitive answers to questions about the number, description, and measurement of religious attitude dimensions.

METHOD

Inverse factor-analytic technique was used as the basic research tool of the present study, which was confined to religious orientations of the Christian culture. Because the completeness of the factor-analytic results depends largely on the adequacy of subject and Q-item coverage of the religious attitude area, the literature was first combed for previously hypothesized religious groupings or clusters and five possible religious types were described. The hypothesized types were as follows: Type 1, stress on sin, judgment, and "thou shalt not's"; Type 2, desire and need for religious belief to fill a void, but lack of specific doctrine; Type 3, stress

on the moral and ethical aspects of religion; Type 4, emphasis on the love and glory of God and worship of Him; and Type 5, "spirit-filled," speaking in tongues, casting out demons, etc. These groupings were seen as clusters within the area of religious attitudes, all of which might be adequately locatable in terms of only one or two factors. They should not be understood as hypothesized factors or dimensions. Q statements designed to represent the religious attitude orientations of these types were then gathered. The preliminary statements were constructed by the author and by persons selected as being representatives of the above types or categories. Additional statements were selected from other attitude or belief scales in the area of "religiosity." These Q statements together with directions for sorting the statements were then submitted to other persons felt to be representatives of the hypothesized categories. These persons were asked to comment on the clarity of the directions for sorting and the readability of the items and to suggest any new statements or changes in the wordings of the old statements. For the final item pool, 133 statements were thus selected.

Four representatives of each of the five hypothesized religious attitude groupings were selected as subjects (Ss). A sixth group composed of two Roman Catholics and two Lutherans picked as being fairly modal members of their respective groups brought the final number of Ss to twenty-four. These Ss were selected from various churches or institutions known to have strong religious orientations in the directions of the hypothesized religious categories. Then, through consultation with administrators of these various institutions, selection was made from the membership of the groups those persons who were seen as being rather strong and "pure" representatives of the desired religious orientations. Table 1 gives the age, sex, education, denominational background, and final rotated factor loadings of each of the Ss.

The Ss were asked to sort the 133 Q statements into nine groups on an agree-disagree continuum. The arrangement of the groups and the number of statements in each group was to be as follows:

Strongly agree 2—7—15—26—33—26—15—7—2
Strongly disagree.

Each of the 24 sorts was then correlated (product-moment correlations) with each of the other sorts and the resulting matrix was factor-analyzed by the centroid method. Because studies such as that of Mosier (4) have shown that more accurate final loadings are obtained if more than the true number of factors are extracted and carried in rotation, the extraction of factors was continued to two factors beyond those indicated to be significant by the application of Humphries' Rule and Tucker's Phi. The axes were then rotated by the single plane method.

The Ss were then divided into two groups for each

¹ This report is based in part on a dissertation submitted to the graduate faculty of the University of Minnesota, March, 1956. The writer wishes to express his appreciation to Professor Paul E. Meehl, Professor Merrill F. Roff, and Robert J. Broen for their generous help in this investigation.

TABLE 1
SUBJECT CLASSIFICATION DATA

Subject	Age	Sex	Education	Denominational Background	Rotated Factor I Loading	Rotated Factor II Loading
Group I						
1.	25	M	2 yrs. college, 1 yr. Bible Coll., 2 yrs. Seminary	Lutheran Brethren	-.08	.70
2.	19	F	3 yrs. HS	Baptist	.02	.65
3.	16	M	3 yrs. HS	Lutheran Brethren	.03	.61
4.	15	F	3 yrs. HS	Lutheran Brethren	.03	.68
Group II						
1.	24	M	BA + 3 yrs. grad. work	Mission Covenant	.25	-.14
2.	38	F	High School	Disciples of Christ	.35	-.02
3.	49	M	7th grade	Disciples of Christ	.48	.33
4.	31	M	BSL and LLB	Disciples of Christ	.20	.01
Group III						
1.	29	F	BA, MA + 3 yrs. additional grad. work	Methodist	.03	-.31
2.	27	M	BA + 3 yrs. grad work	Disciples of Christ	-.06	-.40
3.	64	F	3 yrs. college	Congregational	.48	-.03
4.	20	F	2 yrs. college	Unitarian	-.04	-.19
Group IV						
1.	42	F	2 yrs. college	Disciples of Christ	.54	-.03
2.	62	F	2 yrs. teacher training	Disciples of Christ	.60	-.04
3.	22	F	HS & 3 yrs. nurses training (R.N.)	Disciples of Christ	.71	.04
4.	24	F	BA	Mission Covenant	.22	.04
Group V						
1.	39	M	High school	Pentecostal	.20	.51
2.	44	M	9th grade	Pentecostal	.38	.22
3.	45	F	HS + 1 yr. teacher train., 2 yrs. Bible school, 1 yr. business college	Pentecostal	.40	.20
4.	16	F	2 yrs. HS	Pentecostal	.42	.07
Group VI						
1.	18	F	High school	Roman Catholic	.44	.27
2.	45	F	HS + 1 year beauty school	Augustana Lutheran	.26	.48
3.	46	F	2 yrs. teachers college	Roman Catholic	.42	.32
4.	25	F	HS + 1 yr. Coll. of Medical Technology	Lutheran	.40	.41

RESULTS

final rotated factor (a high loading group and a low loading group), and the average Q sort placement of each Q statement was calculated for each group. The factors were named independently by two judges from a consideration of those statements having the greatest mean differences and least overlap between the high and low loading groups for each factor.

These statements were combined into a Religious Attitude Inventory in an attempt to devise an instrument to measure the factors. The Inventory was of the agree-disagree variety with only those items answered in the plus direction contributing to the factor scores. All statements scored for each factor were weighted equally.

The Religious Attitude Inventory was then subjected to cross-validated study. Directional inventory scale mean differences for different groups were predicted and these predictions then tested by *t* tests.

Application of Humphries' Rule and Tucker's Phi indicated that only two factors were necessary to account for the significant portion of the common variance.

The judges were in almost perfect agreement on the naming of the factors. The first factor (unipolar) was described as "Nearness of God." Persons with high loadings on this factor would tend to feel that God was very real and constantly near and accessible. These persons feel they commune with God—"walk and talk" with Him. Their religiosity consists in emphasizing the Deity's loving presence and guidance rather than His judgment function.

TABLE 2
CROSS-VALIDATIONAL HYPOTHESES AND *t* VALUES

Hypothesis	<i>t</i> value*	<i>df</i>
Factor I		
1. Disciples of Christ over Unitarian	13.19	64
2. Lutheran over Unitarian	14.18	89
3. Theistic Unitarian over Humanistic Unitarian	6.05	42
4. Disciples over Lutheran	5.04	67
5. Lutheran over Theistic Unitarian	8.27	62
Factor II		
1. Lutheran over Unitarian	39.60	89
2. Disciples over Unitarian	9.96	64
3. Lutheran over Disciples	17.91	67
Total Score (I + II)		
1. Lutheran & Disciples over Unitarian	23.20	111
2. Lutheran over Disciples	10.29	67
3. Disciples over Theistic Unitarian	10.76	37
4. Theistic Unitarian over Humanistic Unitarian	4.07	42
Differential Emphasis (I - II)		
1. Unitarian & Disciples over Lutheran	19.97	111

* All tests of hypotheses were significant beyond the .01 level.

The second factor, a bipolar factor correlated +.32 with the first, was described as "Fundamentalism-Humanitarianism." Persons with high positive loadings on this factor would tend to see man as essentially sinful, and emphasize his need for, and rightful fear of, a punishing God. Persons on the other extreme of this factor would tend to see man as containing the potential for good and the tools for actuating this potential (rationality, etc.), thus having little need for much outside intervention in the form of some Deity.

Four groups of persons were used in the Religious Attitude Inventory cross-validation studies: a humanistic Unitarian group; a more theistic Unitarian group; a Disciples of Christ group; and a group from a strict, fundamentalistic Lutheran academy. Besides investigating the validity of Scale I of the Inventory as a measure of Factor I, and Scale II as a measure of Factor II, the cross-validational hypotheses were designed to see if an additive combination of Scales I and II would work as a measure of "general religiosity" (roughly defined as per cent of mental content devoted to religious topics), and if Scale I minus II could be used as a measure of differential (Factor I vs. Factor II) religious emphasis. The cross-validational hypotheses and *t* values for the significance tests of these hypotheses are given in Table 2. The *t* tests were all highly

significant. Because the cross-validational groups were selected and the hypotheses made on the basis of the factor descriptions given above, these data constitute additional evidence concerning the correctness of the factor descriptions.

The cross-validational data presented in Table 2 indicate that Scale I of the Religious Attitude Inventory² can be used as a measure of Factor I, and Scale II as a measure of Factor II. The data also indicate that an additive combination of Scales I and II works very well as a measure of "general religiosity," and Scale I minus II can be used as a measure of differential religious emphasis.

DISCUSSION

In contrast to previous factor-analytic studies which have touched on the area of religious attitudes, this study supports the view that religious attitudes are multidimensional. This difference in findings is probably a function of the somewhat more heterogeneous subjects and test materials used in the present study. It may be that further study of cultures other than the Christian culture investigated in the present study will yield additional factors.

The findings of this study have implications for the interpretation of past research and design of future research in the area of religious attitude correlates. Because the data of this study indicate that there are separate religious attitude factors, it is quite possible that those studies which have investigated the relations between a "general religiosity" and other classes of variables might have shown markedly different results if that "general religiosity" were to have been broken down into its more basic dimensions. Because these separate factors in religious attitude may well have differing and sometimes opposing correlates which are lost when they are combined into an impure variable of "general religiosity," measures of the separate religious attitude factors should

² The original correlation matrix, the unrotated factor matrix, and the Religious Attitude Inventory with keys for scoring have been deposited with the American Documentation Institute. Order Document No. 5066 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance. \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

be used in investigations concerned with the correlates of religious attitudes. The Religious Attitude Inventory here developed appears to provide fairly valid scales for the measurement of the factors found in the present investigation.

SUMMARY

Two basic religious attitude dimensions were isolated by inverse factor analysis. The first was described as a "Nearness of God" dimension and the other as a "Fundamentalism-Humanitarianism" dimension. A Religious Attitude Inventory with scales designed to measure these factors was constructed and subjected to cross-validated study.

REFERENCES

1. CARLSON, H. B. Attitudes of undergraduate students. *J. soc. Psychol.* 1934, **5**, 202-213.
2. FERGUSON, L. W. A revision of the primary social attitude scales. *J. Psychol.* 1944, **17**, 229-241.
3. JOHNSON, MURIAL. A preliminary study of social attitudes. Unpublished bachelor's dissertation, University Coll., London, 1941.
4. MOSIER, C. I. Influence of chance error on simple structure: an empirical investigation of the effect of chance error and estimated communalities on simple structure in factorial analysis. *Psychometrika*, 1939, **4**, 33-44.
5. SANAI, M. An empirical study of political, religious, and social attitudes. *Brit. J. Psychol., Statist. Sect.* 1952, **5** (Part II), 81-92.
6. THURSTONE, L. L. The vectors of mind. *Psychol. Rev.*, 1934, **41**, 1-32.

Received April 16, 1956.

OPINION CONGRUENCE WITH A NEGATIVE SOURCE ON ONE ISSUE AS A FACTOR INFLUENCING AGREEMENT ON ANOTHER ISSUE¹

WALTER WEISS

Boston University

IN the "natural" communication setting, people are often exposed to more than one communication by a known source. The contents of the several communications may be focused on the same or related issues or may refer to unrelated topics, which may or may not be opinion issues. Under these circumstances, audience response to any one communication may be affected by the kinds of responses made to prior communications by the same source. A specific, interactive relationship suggested by Freeman (1), which is encompassed by this general assumption, was made the focus of the current research. In an impressionistic analysis of the procedures and techniques of propagandists, he noted that such communicators often interlard or precede the primary persuasive communication with assertions in favor of one or more of the norms, mores, or morals of the audience being addressed. For example, if the primary intent of a speaker was to convince a prohibitionist group to adopt a certain position on an issue of foreign affairs, he might begin his address with statements that bemoaned sin and the curse of liquor. Freeman called this propaganda technique "flogging the dead horse." It was presumed by him that such assertions of agreement with the communicatees on a matter of importance to them would enhance their receptivity to a following, persuasive message. The main purpose of the current research was to determine the relative efficacy of this procedure. Although a credible source could benefit from the prior establishment of such opinion congruence, demonstration of the effect seemed more likely with a noncredible communicator.

The research design was elaborated to provide some information on two other, quite different factors. First, for pragmatic and theoretical reasons, it seemed worth while to

examine the possibility of a "boomerang" effect as a result of a statement of opinion congruence by a distinctly negative source. That is, the communication of opinion congruence by a poorly regarded source might induce an unfavorable opinion change on the topic, resulting in reduced congruity between the opinions of the communicatees and those of the source. Secondly, in discussions of the communicator as a factor in the communication process, it has been implicitly assumed that the effective communicator is the "external" one, to whom the authorship of the total message is attributed. In a typical communication however, there are often one or more "internal" communicators, with the same or different credibility ratings as the external one, that are used as sources for relevant data and assertions. With respect to communication effectiveness, an untrustworthy external source might increase the opinion impact of his message by the judicious use of more prestigious, internal sources. The relevant comparisons to assess this possibility were also included in the research design.

METHOD

Experimental design. A before-after design was employed, with a class of approximately 120 students at Boston University comprising the sample. The before-Boston University was presented as an opinion survey on a variety of topics, such as academic freedom, fluoridation, juvenile delinquency, discrimination against minority groups, the internal security system, and the effects of mass media on children. Opinions on each of these topics were assessed by a set of four questions. The students also indicated the general importance to them of the opinion topics by rating each on an 11-point scale. Finally, the Ss rated the general trustworthiness, as sources of information, of a variety of people, magazines, and newspapers on a 0-to-100 basis.

From the responses to the questionnaire, two topics and two sources were selected as appropriate for the requirements of the research. Since the mean position of the Ss was favorable to the concept of academic freedom and since it was an area of importance to them, this opinion issue was the one chosen for "flogging." The topic of fluoridation was selected as the focus for the persuasive communication; most Ss were in favor of

¹ The research reported in this article was done under Contract NONR-492(04), between Boston University and the Office of Naval Research.

fluoridation but did not rate the topic particularly high in importance. Almost uniform agreement on source was found only for the two chosen: the *New York Daily Worker*, as a highly untrustworthy source, and the *New York Times*, as a trustworthy one.

Two articles opposing fluoridation were prepared. One, hereafter designated as fluoridation (A), contained evidence and quotations supporting the antifluoridation position, that were attributed to certain named governmental and professional sources. The other, hereafter designated as fluoridation (B), contained the exact same content, but with all quotation marks and references to other sources omitted. Thus, for the people reading fluoridation (B), there was no attributed source other than the external one for any of the information in the article.

The communication session occurred approximately eight weeks later. Four experimental groups were formed. Group 1 read an article in favor of academic freedom immediately followed by fluoridation (A), both with the *Daily Worker* as source. Group 2 read a completely descriptive, nonpolitical article on the territory of Uganda in central Africa, followed by fluoridation (A), again with the *Daily Worker* as the source for both. Group 3 read the article on academic freedom with the *New York Times* as source and then fluoridation (A) with the *Daily Worker* as source. Group 4 read the article on Uganda and then fluoridation (B), with the *Daily Worker* as the source for both.

The students were informed that the research was concerned with people's reactions to newspaper articles and that immediately following the reading, for which 10 minutes would be allowed, they would be asked a series of questions on their reactions to the articles and on the content of the articles. At the top of the title page of each article, one of the following statements was made: "This article was taken from a recent issue of the *New York Times*;" or "This article was taken from a recent issue of the *New York Daily Worker*, the official newspaper of the Communist Party in the United States." The articles on fluoridation and academic freedom were well-organized, reasoned statements of position on these topics; their tone was calm and forceful, not shrill or aggressive.

The four pairs of articles, corresponding to the treatment combinations mentioned above, were distributed as booklets in a prepared random order, with somewhat more Ss being assigned to the first and second groups than to the remaining ones. The students' names were not requested; instead, the before-and-after questionnaires were matched on the basis of responses to a set of personal information items.

Immediately after reading the articles, the Ss responded to a questionnaire that contained a series of items on their judgments of the articles, on their opinions on the topic of fluoridation, and on the content of the article on fluoridation. In answering each judgmental question, a response concerning the fluoridation article was called for prior to one concerning the other article the student had read. For Groups 1 and 3, the same four opinion questions on academic freedom contained in the before-questionnaires were also included in their after-questionnaires. These followed in order the questions on fluoridation. Six opinion items on fluoridation were employed in all after-question-

naires; three of these had been responded to in the before-questionnaire and three were new ones. One before-item was not included in the after-questionnaire, since its focus was neither directly nor indirectly dealt with in the article on fluoridation.

Opinion scoring. Five response alternatives were available for each opinion question: strongly agree, agree, uncertain, disagree, strongly disagree. In scoring the items, no distinction was made between the two "agree" and between the two "disagree" categories. When a response category indicating approval of fluoridation or academic freedom was selected, a score of +1 was assigned; when disapproval was indicated, a score of -1 was given. A response of uncertainty was scored 0. Thus, the theoretical ranges were from +4 to -4 for the total before-opinions on each topic and for after-opinions on academic freedom, and from +6 to -6 for the after-opinions on fluoridation.

RESULTS

The mean opinions on fluoridation before and after exposure to the communication for each experimental group are given in Table 1. Although there are no significant differences among the means at either time, the difference in after-opinions between Group 1 (the group receiving the "flogging" treatment) and the combined Groups 2 and 3 (the ones exposed to the same fluoridation article) is in the predicted direction ($p = .10$, one tail); unfortunately, variability of response within the groups is quite high, thus affecting the reliability of the mean differences. (Employing the before-means in a covariance analysis to adjust the after-scores yielded similar results.) If only the sign of the total after-score is considered, then the percentages of Ss in each group whose total scores are in the minus category (i.e., indicate in general an antifluoridation position) can be compared. The relevant data are presented in Table 2. The differences between Groups 1 and 2 and between Groups 1 and 3 are significant ($p < .05$, one tail, for each); none of the other comparisons approaches significance.

A prerequisite condition for the effectiveness

TABLE 1
MEAN OPINION ON FLUORIDATION BEFORE AND AFTER
EXPOSURE TO COMMUNICATION

Time of Questioning	Group 1	Group 2	Group 3	Group 4
Before opinion*	+2.31	+1.79	+1.86	+1.96
After opinion**	-0.44	+0.28	+0.59	-0.44
	36	39	22	25

* Based on four items.

** Based on six items.

TABLE 2

PERCENTAGE OF Ss IN EACH GROUP WHOSE AFTER OPINIONS ON FLUORIDATION WERE UNFAVORABLE (Based on six items)

	Group 1	Group 2	Group 3	Group 4
Percentage	63.9	39.5	36.4	48.0
N	36	39	22	25

TABLE 3

RELATION BETWEEN AFTER OPINIONS ON ACADEMIC FREEDOM AND OPINIONS ON FLUORIDATION

Fluoridation	Academic Freedom			
	Favorable*		Not favorable**	
	Before	After	Before	After
Mean	+2.22	-1.09	+2.46	+0.69
N	23	23	13	13

* Opinion score was plus.

** Opinion score was minus.

of the "flogging" procedure is the existence of opinion congruence on the first topic. Since only two thirds of the Ss in Group 1 exhibited after-opinions that were favorable toward academic freedom (i.e., total opinion scores that were in the plus category), a more precise test of the hypothesis requires that the opinions of these Ss on fluoridation be considered separately from those of the Ss in this group who were unfavorable toward academic freedom. The relevant means are given in Table 3. The difference between the before-means of these subgroups is not significant; but the difference between the after-means is significant at the .05 level, one tail. Furthermore, the after-mean of the "congruent" subgroup is significantly different in the predicted direction from those for Groups 2 and 3, as given in Table 1 ($p = .05$, one tail for both differences); this is not the case for the "noncongruent" subgroup which is slightly, though not reliably, more profluoridation than Groups 2 and 3.²

For the total group of students used in this research, there was no significant correlation between initial opinions on fluoridation and on academic freedom. Also, there are no significant

² The same results are obtained whether the subgroups are formed on the basis of before-opinions or unchanged before-after opinions. Although some Ss exhibited minor shifts in opinion position, very few changed from the designated favorable to the unfavorable category or the reverse.

TABLE 4

AFTER OPINIONS ON FLUORIDATION RELATED TO CERTAIN JUDGMENTS OF FLUORIDATION ARTICLE

	Learned a Lot		Other	
	Group 1	Groups 2 & 3	Group 1	Groups 2 & 3
Mean	-0.43	-1.50	-0.32	+1.25
N	14	20	22	40
p (two tails)	.20		.09	
	Mostly True		Other	
	Group 1	Groups 2 & 3	Group 1	Groups 2 & 3
Mean	-0.50	-1.35	-0.42	+1.00
N	10	17	26	43
p (two tails)	> .50		.06	

differences among the four experimental groups on the fact-quiz test of knowledge of the content of the article on fluoridation nor on the mean ratings of the importance of this topic to the students. The grand mean on the fact-quiz test was 4.0 (with 5.0 the maximum possible); and the over-all mean rating of importance was 4.6 (with 11.0 the maximum possible).

To obtain additional information on the effects of the "flogging" procedure, the relationship between certain judgmental responses and after-opinions on fluoridation was examined. Since most of the questions yielded extremely skewed distributions, only two of the judgments could be utilized in this analysis. These concerned the S's estimation of how much he learned from the article on fluoridation and how true he thought its content was. The relevant data are contained in Table 4.³ There were no significant differences in before-opinions or in fact-quiz scores among these subgroups. The effectiveness of the "flogging" procedure in facilitating opinion change is exhibited by the subgroups of Ss who did not judge the article as mostly true or who responded that they did not learn a lot from the article. For those Ss who indicated that they learned a lot or that the article contained mostly true information, the difference between the subgroups, though not significant, is in the opposite direction. Within Group 1

³ The data for Groups 2 and 3 are combined, since their mean opinion scores did not differ significantly from each other. Comparisons between each of these groups and Group 1 yield results comparable to those obtained with the data as presented in Table 4.

TABLE 5
MEAN BEFORE AND AFTER OPINIONS
ON ACADEMIC FREEDOM
(Based on four items)

	Before		After	
	Group 1	Group 3	Group 1	Group 3
Mean	+1.44	+1.50	+1.31	+1.45
N	36	22	36	22
p (two tails)	> .90		> .80	

TABLE 6
PERCENTAGE OF Ss IN EACH GROUP INDICATING
CERTAIN JUDGMENTS OF THE ARTICLES
ON ACADEMIC FREEDOM AND UGANDA

	Academic Freedom		Uganda
	<i>Daily Worker</i>	<i>Times</i>	<i>Daily Worker</i>
Mostly one-sided	41.7	9.1	15.9
Yes, propaganda	58.3	50.0	27.0
Mostly true	41.7	90.9	61.9
N	36	22	63

judgments of the article were not related to the extent of opinion change; but the differences between the judgment subgroups within the combined Groups 2 and 3 were significant for each question ($p < .03$, two tails, for both comparisons).

Table 5 contains the mean opinion scores on academic freedom of the two groups given the same opinion congruent communications but with different attributed communicators: the *New York Daily Worker* for one group and the *New York Times* for the other. The groups do not differ significantly from each other in their before or after opinions.

There were no differences among the groups in the Ss' judgments of the article on fluoridation. Most Ss considered it one-sided or a piece of propaganda but indicated that they learned something from the article; few judged its content as mostly true. An analysis was also made of the Ss' judgments of the first article they read. The relevant data are presented in Table 6.⁴ Those Ss reading the article on academic freedom with the *Daily Worker* as source were more likely to judge it as mostly one-sided ($p = .02$) and less likely to consider its

content as mostly true ($p < .001$) than were those reading it with the *New York Times* as the attributed source; but these groups did not differ significantly in their judgments of the article as a piece of propaganda. Comparisons between the groups reading the academic freedom article with the *Daily Worker* as source and the article on Uganda with the same source reveal that the latter Ss were less likely to judge their article as mostly one-sided ($p < .01$) or a piece of propaganda ($p < .01$) and somewhat more likely to judge it as containing mostly true information ($p = .09$) than were the former Ss. Furthermore, the Ss reading the Uganda article with the *Daily Worker* as source were somewhat less likely to judge it as propaganda or as mostly true than were those Ss reading the article on academic freedom with the *New York Times* as source ($p = .08$ and $.02$, respectively); but these groups did not differ significantly in their judgments of the articles as one-sided presentations.

DISCUSSION

The research findings offer support for the hypothesis that a statement by a communicator of opinion congruence on a topic of importance to the communicatees facilitates the opinion-change effectiveness of a following, persuasive communication. However, the nature of the facilitating psychological processes induced by the "flogging" procedure have still to be determined. It is possible that an enhancement of the trustworthiness of the source may be a factor, although with the particular source used this is not likely to have occurred. Also, judgments of the persuasive communication (the article on fluoridation) were apparently not affected differentially by the experimental conditions.⁵ Opinion congruence rather than simply a factual, objective presentation on a noncontroversial matter seems to be the critical factor, even though the latter kind of communication (the article on Uganda) provokes more favorable judgments by the audience.

It may be that the establishment of congruence or agreement with an audience on one topic inhibits the arousal of interfering re-

⁴ The data for the two groups exposed to the article on Uganda are combined, since the differences between them in their judgments of the article did not approach significance. Also, the p values given in the text are all two-tailed ones.

⁵ It should be noted that the current data as well as previous research (2, 8, 9) confirm a general relationship between certain judgments and the effectiveness of a communication.

sponses *during* exposure to the following, persuasive message, thus permitting it to have a greater impact. The importance of implicit interfering responses as factors in communication effectiveness has been noted by other researchers (2, 5, 6). In the current study these responses might have taken the form of recall of countering data and arguments, lack of careful consideration of the significance and relevance of the presented facts and statements, or arousal of negative judgments of the validity of the content while the article was being read. It should not be inferred from this supposition, however, that reduction or inhibition of interfering responses results in the automatic acceptance of persuasive arguments; what is being asserted is that certain resistances to acceptance may be restrained, thereby permitting the communication to have its more "natural" impact. One implication of the assumed significance of interfering responses in the general interactive relationship between communications effects is that an initial message which presents a viewpoint opposed to one strongly held by an audience may reduce the effectiveness of a second persuasive communication on another issue of lesser centrality for the audience. That is, the predisposition to make implicit interfering responses during exposure to the second communication may be activated by the initial, provoking appeal.

As potential, qualifying influences on the general effectiveness of the "flogging" procedure, several factors connected with the current research should be mentioned. These are source credibility, the importance of the topics to the audience, the lapse of time between the two communications, and the audience's interpretation of the communicator's assertions on the first topic. Although the establishment of opinion congruence may enhance the opinion influence of a highly regarded communicator, it is not likely to have as great an effect as with an untrustworthy one. This prediction is based on the supposition that source credibility as a critical judgment by the audience is negatively related to the arousal of interfering responses. Thus, the contribution of initial opinion congruence as a further inhibiting influence may not be significant. With respect to the importance of the topics of the communications, the particular

issues employed in this study were selected so as to provide the best judged setting for the demonstration of the effectiveness of the critical experimental procedure. Further research will be needed to examine the influence of initial opinion congruence when the topic of the first communicator is of little importance or when the topic of the persuasive communication is of great importance to the communicatees. Also, the effectiveness of this procedure may be found to vary inversely with the extent of delay between exposure to the first and second communications; in this study, delay was minimal. Finally, the predispositional effectiveness of the communicator's statement of opinion congruence may be greatly modified by the audience's judgments of it as insincere, specially motivated, or inconsistent with the known, past behavior of the source. Unfortunately, data relevant to such judgments were not obtained in this research.

Although the experimental treatments did not affect differentially the Ss' judgments of the article on fluoridation, an apparent qualifying relationship was found between certain judgments and opinion change on this topic (see Table 4). The form of the results suggests that Ss who are not exposed to the special, opinion congruent treatment, but for whom the antifuoridation arguments are sufficiently impressive to overcome interfering responses and change-resisting influences, exhibit the largest effect of the communication. This is analogous to the behavior, observed in everyday experience, of a person who is initially quite resistant to some interpretation or judgment but who ends up by adopting it more strongly and extremely than others, after surrendering to the great appeal the viewpoint has for him or being thoroughly convinced of the validity of the new position. A frequent accompaniment of this result is the firm rejection of judgments and previously held opinions inconsistent with the adopted viewpoint.

The assertion has sometimes been made that popular support for a particular idea or opinion position can be radically diminished by the actual or imputed adherence to this viewpoint of an especially untrustworthy and suspect communicator. Some of the data of the current research do not confirm the general validity of this broad, unanalyzed contention; for no differential regression to a less favorable posi-

tion on academic freedom occurred following statements of opinion congruence by the *Daily Worker* and the *New York Times*.⁶ It would seem therefore that the hypothesized consequence, if actually valid, may be functionally dependent on the influence of a compound of antecedent conditions.

In this research, "internal" sources who could be considered more credible than the main "external" communicator did not enhance the measured effectiveness of the fluoridation article. Although the students' judgments of the general trustworthiness of doctors, dentists, scientists, and congressmen were not determined, it seems reasonable to suppose that such sources would be rated more highly than the *Daily Worker*, especially on the topic of the article. However, the particular internal sources named were not familiar to the students. Consequently, the general context of low credibility evoked by the primary communicator may have induced judgments of associated untrustworthiness with respect to the internal sources. That is, the communicatees may have considered these sources as biased pleaders for the anti-fluoridation point of view, since they were selected by the *Daily Worker* to support its one-sided presentation.

Previous research has indicated that the audience's assumptions concerning the trustworthiness or intent of the communicator influences certain judgments of the communication itself (3, 4, 7). In this study, some data are at hand bearing on the relative contributions of source credibility and the kind of content communicated to such judgments. In terms of the articles on academic freedom and Uganda, the factor of content differentiation may be considered as reflecting the combined dimensions of relatedness to an opinion issue and of persuasion toward adopting the communicator's viewpoint. Thus, the academic freedom article may be judged an instance of

an explicit attempt at persuasion on an opinion issue; and the one on Uganda can be viewed as a nonpersuasive, noncontroversial article. The relevant results (see Table 6) seem to indicate that communicatees judge attempted opinion persuasion, despite differences in communicator trustworthiness, as propaganda; this supports a supposition expressed by Weiss and Fine (9). In other words, as determinants of this judgment, the content factor appears to be more influential than that of source credibility. However, the judgments of the articles as mostly one-sided or mostly true seem to be influenced considerably by either of the two factors, with both contributing comparably to the former judgment and with source somewhat more of a determinant for the latter one.

SUMMARY

The primary purpose of the research was to test the hypothesis that a statement by a communicator of opinion congruence with the communicatees on an issue of importance to them will facilitate the opinion-change effectiveness of a following, persuasive communication on a different topic. Subsidiary questions concerned: Would a statement of opinion congruence by a particularly untrustworthy and suspect source produce a "boomerang" effect? And, is the effectiveness of a communication by a negative source enhanced by the attribution of significant supporting data and arguments to more trustworthy, "internal" sources?

A before-after design was employed, with approximately 120 college students as Ss. The communicated statement of opinion congruence expressed a pro-academic freedom position; and the persuasive communication argued for an anti-fluoridation position. In the communication session, approximately eight weeks after their opinions on these issues had been ascertained, four experimental groups were formed. Group 1 read an article in favor of academic freedom immediately followed by one opposing fluoridation, both attributed to the *Daily Worker*. Group 2 read an article on Uganda, followed by the one opposing fluoridation, again with the *Daily Worker* as the source for both. Group 3 read the article on academic freedom with the *New York Times* as source and then the one on fluoridation with the *Daily Worker* as source. Group 4 read

⁶ In terms of the particular experimental design employed, this interpretation of the relevant results depends on the assumption that the credible source itself did not induce a "boomerang" effect. This is not likely however in view of the often confirmed finding that trustworthy sources exert a strong, positive influence on opinions (3, 5, 7). In a more precise design, before-after opinions on academic freedom would have been obtained from a standard control group; certain practical, methodological considerations made this arrangement difficult to employ.

articles on Uganda and on fluoridation, with the *Daily Worker* as the source for both. However, for these Ss, all quotation marks and references to other internal sources were eliminated from their article on fluoridation. All Ss responded to an immediate-after questionnaire. The following results were noted:

1. The data provide support for the main research hypothesis concerning the facilitating influence of the prior establishment of opinion congruence on the effectiveness of a following, persuasive message. There were no differences among the groups in their judgments of the opinion-change communication as fair, propagandistic, or true.

2. An apparent, qualifying effect on opinion change was observed between experimental conditions and certain judgments of the persuasive communication. Judgment was unrelated to opinion change for Group 1 but was related for Groups 2 and 3; furthermore, those in Groups 2 or 3 who expressed favorable judgments exhibited the largest opinion change among the judgment-procedure subgroups.

3. The communication of opinion congruence on a particular issue by an untrustworthy source did not lead to a "boomerang" effect on the communicatees' opinions on this issue.

4. On the basis of the current procedures, there was no apparent enhancement of communication effectiveness by attributing significant aspects of the content of the persuasive communication to internal sources more trustworthy than the primary, untrustworthy one.

5. Judgment of a communication as propagandistic seems to be related primarily to apparent intent to persuade on an opinion topic rather than to source credibility. But judgments of fairness and truthfulness are affected by both factors.

REFERENCES

1. FREEMAN, E. *Conquering the man in the street*. New York: The Vanguard Press, 1940.
2. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communications*. Princeton: Princeton Univer. Press, 1949.
3. HOVLAND, C. I., & WEISS, W. The influence of source credibility on communication effectiveness. *Publ. Opin. Quart.*, 1951, 15, 635-650.
4. HOVLAND, C. I., & MANDELL, W. An experimental comparison of conclusion-drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.*, 1952, 47, 581-588.
5. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.
6. KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, 6, 185-214.
7. KELMAN, H. C., & HOVLAND, C. I. "Reinstatement" of the communicator in delayed measurement of opinion change. *J. abnorm. soc. Psychol.*, 1953, 48, 327-335.
8. WEISS, W., & FINE, B. J. Opinion change as a function of some intrapersonal attributes of the communicatees. *J. abnorm. soc. Psychol.*, 1955, 51, 246-253.
9. WEISS, W., & FINE, B. J. The effect of induced aggression on opinion change. *J. abnorm. soc. Psychol.*, 1956, 52, 109-114.

Received April 19, 1956

EXTINCTION DURING PSYCHOTHERAPY OF GSR ACCOMPANYING "EMBARRASSING" STATEMENTS¹

JAMES E. DITTES

Yale University

ALTHOUGH changes in the electrical conductivity of the palmar surfaces of the hand are known to accompany emotional responses in general (2, 5) and palmar sweating has been found correlated with other measures of general tension during psychotherapy (4), little is known about which, if any, specific emotions or attitudes are accompanied during therapy by measurable galvanic skin responses. Unless the variables which it supposedly measures can be more precisely specified, the GSR is unlikely to prove useful in research on psychotherapeutic problems.

An emotion with which psychotherapy is centrally concerned is fear, especially when associated with patient reports of certain thoughts or behavior which have been historically punished, particularly by the parents. Because such thoughts or behavior are reported by the patient to the therapist without ensuing punishment, the associated fear should extinguish. Extinction also should generalize to any existing fear which may be evoked by simply thinking such thoughts, as well as that evoked by reporting them (1).

One type of thought commonly supposed to arouse such transferred fear is a report of personal sexual wishes or activity. This study attempts to discover whether the GSR can be used to trace the arousal and extinction of fear connected with making such "embarrassing" sex statements to the therapist.

Two interpretations might be offered for the coincidence of a galvanic skin response and a statement describing sexual behavior or wishes. One is that the emotion indicated by the GSR

is the fear of making the statement to another person or the fear evoked by acknowledging the sexual impulse or wishes to oneself. Another explanation is that the GSR represents the "re-experienced" emotion which is reported, that sexual feelings themselves are aroused in the reporting and that this sexual mobilization produces the GSR. However, this second interpretation would not predict the extinction of the GSR during therapy, since extraclinical situations are presumably continuing to reinforce the evocation of sexual mobilization in response to sexual thoughts as much during the period of therapy as previously. A finding, then, that statements reporting sexual behavior and wishes are not accompanied by GSR late in therapy as frequently as in early hours affords some basis for determining whether the GSR reflects a particular emotion indicated in the contents of a patient's remarks or the feelings of fear associated with making the report. The hypothesis proposed here is that the GSR more readily measures the latter.

METHOD

This study is based on the records of a 35-year-old, white, male, neurotic patient seen by an experienced therapist weekly for an hour and a half for 30 sessions. All sessions were electrically recorded, and verbatim transcripts were typed.

A continuous record of the patient's palmar skin conductance was made during each session after the first two (except for hours 18 and 22, during which the equipment failed) with the following apparatus:

1. Small electrodes in a Bakelite holder were taped to the palmar surface of the patient's finger. The electrodes were made of analytic reagent silver .001 inch thick and approximately 1 cm. wide and 2 cm. long. An electrode paste made of bentonite, glycerine, and isotonic saline solution was used.

2. A dry-cell voltage source, including a variable voltage so that the voltage drop across the patient's basic resistance level can be balanced out, allowed an amplifier to be used in its most sensitive range and also provided a measure of the basic level.

3. A Sanborn model 126 DC amplifier was used.

4. A Sanborn model 127 galvanic recorder recorded changes in conductance on a continuously moving, calibrated paper.

5. A device was employed which periodically produces characteristic pips on the recording paper simul-

¹ This study was done as part of a psychotherapy research program supervised by Dr. John Dollard and supported by United States Public Health Service Grant M-648. Appreciation is due Dr. Frank Auld, Jr., and Dr. Dollard for their suggestions and assistance. The data used here had all been previously obtained by them in other phases of the research program. Mr. Harry W. Dreyer, electronics engineering consultant to the project, designed and built some of the apparatus used. Mrs. Ann C. Byler served as a judge of the verbal material.

TABLE 1

COMPUTATION OF PHI COEFFICIENT OF RELIABILITY OF ESS JUDGMENTS FOR HOUR 15

		Judge A	
		Sex speeches judged as ESS	Sex speeches not judged as ESS
Judge B	Sex speeches judged as ESS	6	1
	Sex speeches not judged as ESS	1	21
Total number of sex speeches as selected by Judges C and D		29	
Phi coefficient of reliability		.81	

taneously with beeps on the sound tape. This device could also be manually operated to produce an identifying pattern of pips and beeps at the beginning and any other point to insure absolute synchrony for subsequent playback.

Judging embarrassing sex statements. The typed transcript of each session was read independently by two judges, who picked out statements judged to be embarrassing sex statements (subsequently abbreviated ESS), according to a definition summarized as follows: An ESS is (a) a detailed description of a specific occasion of personal sexual behavior, including details of the patient's own or his partner's body or mention of orgasm and implying his responsibility for the event at least in part; or (b) a positive and emphatic statement of strong sex desires, apart from any particular episode; or (c) use of vulgar sex vocabulary. A statement was considered to be an ESS if both judges independently selected it. One judge was ignorant of the specific hypothesis being tested. Hours were judged in random order, except that all the odd-numbered hours were judged before the even-numbered.

The unit of a "statement" was not defined more objectively than as a verbalization of a single thought or topic. However, the reliability of the implicit units used may be assessed in part by noting that each judge indicated the line of the typescript in which the statement was judged to begin and that the average difference between the judges, for those statements which both selected, was .48 of a line. They selected the same line of typing in well over half of the cases and differed by only one line on most of the others.

A phi coefficient of correlation was used to assess the reliability of the judging of ESSs. There were available, as an estimate of the population of sex statements from which the ESSs were selected, judgments of sex content which had been made by other judges in connection with another study using the same material.² Table 1 illustrates, for Hour 15, the median

² For purposes of obtaining this reliability estimate, a unit of a "speech" was objectively defined as all the remarks of the patient between statements by the therapist or between pauses of at least five seconds, or else ten typed lines if the speech was otherwise more than that. Since this unit of a "speech" is larger and there-

TABLE 2

NUMBER OF ESS AND RELIABILITY COEFFICIENTS OF ESS JUDGMENTS FOR EACH HOUR

Hour*	Number of Embarrassing Sex Statements			Phi Coefficient
	Judge A	Judge B	Agreed	
3	1	1	1	1.00
4	1	2	1	.67
5	1	1	0	—
7	14	13	7	.41
8	13	13	10	.63
9	14	15	10	.56
10	8	8	7	.82
11	20	21	16	.52
13	11	10	7	.49
15	7	7	6	.81
16	8	10	5	.44
17	4	5	4	.85
19	14	15	14	.95
20	25	26	23	.86
21	5	4	4	.88
25	7	5	5	.81
29	6	9	6	.74
30	7	5	5	.63

* No ESS was found by either judge for hours 6, 12, 14, 23, 24, 26, 27, or 28, which are omitted from this table.

hour, how the phi coefficient was calculated, based on the number of those statements selected as an ESS by both judges, and the number selected by only one or the other of the judges, from the total population of sex speeches, as determined independently by other judges. Table 2 reports, for each hour, the phi coefficient and the actual number of ESS judgments.

Scoring of GSR record. A GSR was scored whenever the continuous record of conductance met one of the following criteria: (a) a rise at the rate of 4,000 or more ohms per second with an amplitude of at least 6,500 ohms; or (b) a rise of at least 10,000 ohms amplitude at a rate of 1,333 or more ohms per second. Such a rise represented a change of the order of 3 per cent of the basal level. These criteria were selected after an independent study of the GSR record indicated that they differentiated fairly clearly between clear-cut abrupt rises in conductance and smaller fluctuations. There were few borderline observations close to the criteria. These values were easily distinguishable with the calibrated paper used.

Despite indications (5) that the GSR actually varies in proportion with the basic resistance level, the criteria for the GSR were kept standard for all sessions because the variation of basic resistance level was not great. The mean basic level for the start of the hour was 360,400 ohms, with the particular apparatus used, and the series of hours varied with a standard deviation of 88,000 ohms, less than one-fourth the mean.

RESULTS

The solid line in Fig. 1 shows for each therapy session the percentage of ESS on fore less numerous than the unit of a "statement" used in the actual judging, the reliability coefficient is biased in a conservative direction.

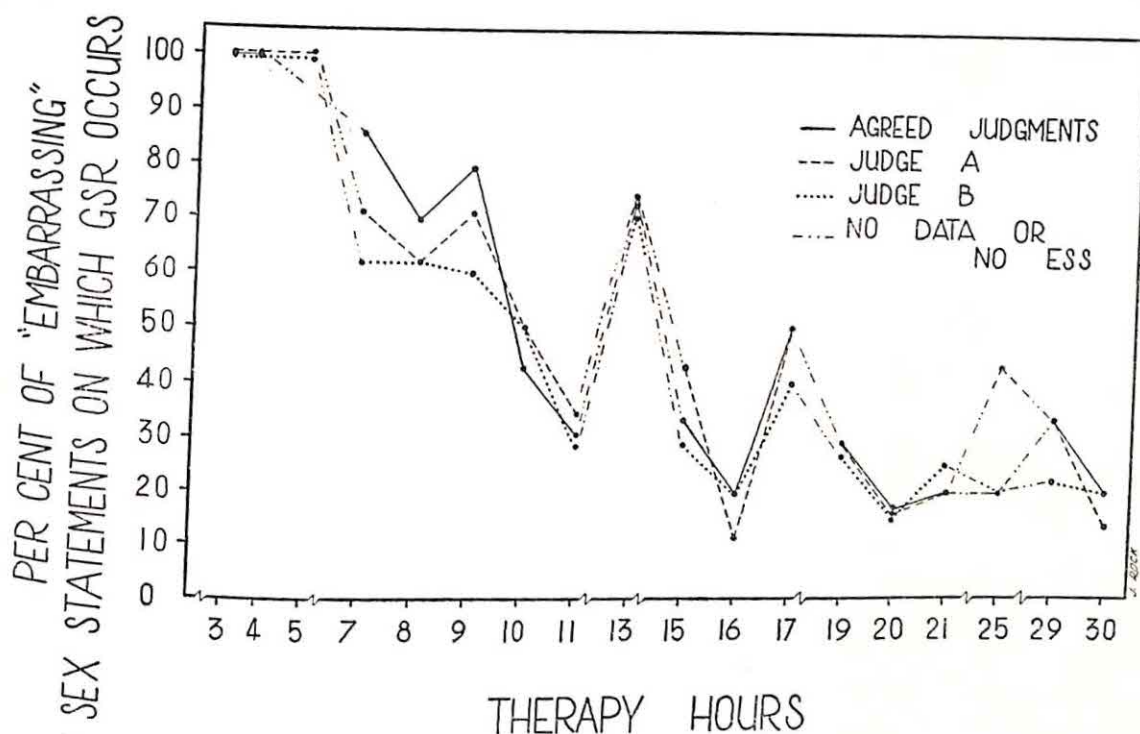


FIG. 1. GSR ACCOMPANYING "EMBARRASSING" SEX STATEMENTS IN SUCCESSIVE THERAPY HOURS
No GSR data are available for Hours 1, 2, 18, and 22. No ESS was found by judges for Hours 6, 12, 14, 23, 24, 26, 27, and 28.

which a GSR rise occurred. No data are given for hours 5, 6, 12, 14, 23, 24, 26, 27, and 28, during which no ESS was found. To test the decline in percentage of GSR from early to late hours, a division into four quadrants at the median hour and median percentage yielded a chi square of 7.42, significant beyond the .01 level. Estimation of the correlation between sequence of hours and the GSR percentage resulted in a rho of .85 and a tau of .71, significant beyond the .0001 level.³

Various types of analysis were made in order to rule out possible alternative explanations of the results. To test the possibility that the results could be a consequence of biases in judging ESS by the one judge who was aware of the hypothesis, the percentage of rises on statements selected by each judge are shown by the broken lines in Fig. 1. The trend is similar whether the agreed upon or the separate judgments are used.

To test the possibility that the lack of an objective definition for the unit of statement

allowed for bias in determining whether or not a GSR occurred during a particular statement, three analyses were made. First, results were calculated using as an arbitrary objectively defined unit a "speech" of the patient, defined as all his remarks occurring between remarks by the therapist or between pauses of at least five seconds, or ten typed lines if the speech was otherwise longer than that. Such a "speech" would ordinarily include several "statements." Using any GSR which occurred within the same speech as an ESS, the results still show an over-all decline. With a division at the median hour and the median percentage, chi square was 4.58, significant at beyond the .05 level.

Second, it was noted that the average discrepancy, in typed lines, between the beginning of an ESS as determined, on the average, by the judges and the beginning of the GSR was .8. No GSR which occurred within one line of the judged beginning of an ESS was excluded. This figure of .8 can be compared with the average of 4.3 lines discrepancy between ESS and the nearest nonscored GSR occurring within the same "speech" as previously defined.

³ Although generally yielding a lower absolute figure than rho for comparable data, tau is used in the analysis because it permits partial correlation and a test of significance involving fewer assumptions (3).

TABLE 3
RANK-ORDER INTERCORRELATIONS AMONG VARIABLES*

	Percent- age of GSR on ESS	Total Number of ESS	Total Number of GSR	Number of Sex Speeches
Sequence	.71	-.05	.19	.02
Percentage of GSR on ESS		-.21	.17	-.24
Total Number of ESS			-.09	.48
Total number of GSR				-.09

* Tau coefficient used.

Third, the average discrepancy in typed lines between GSR and ESS was .86 for the first half of the series of hours and .79 for the second half, a difference hardly great enough to suggest any systematic bias in scoring the correspondence of GSR and ESS.

Table 3 presents tau coefficients of rank correlation among variables which might be suspected of producing an artifactual relationship between sequence of hours and GSR percentage of ESS. None of these relationships approaches significance except the predicted one and that between number of ESS and number of sex speeches, which is hardly unexpected and which suggests a certain utility in the judgments of ESS. The partial rank correlation of sequence and GSR percentage with number of ESS partialled out is .70, and the same correlation with number of GSR partialled out is .73.

DISCUSSION

The results support the hypothesis proposed on the basis of the two assumptions that the GSR is associated with feelings concerning the immediate interpersonal relationship and that such feelings should extinguish in the course of therapy. The GSR can be interpreted as an index of transferred embarrassment over reporting sexual material in therapy. Whether the GSR could be similarly used to indicate other feelings towards the therapist remains to be determined. It seems likely that most aspects of "negative" transference would show the same kind of relationship with GSR as that reported here.

Study of the three serious exceptions to the general declining trend shown in Fig. 1 offers incidental support to the argument that the

GSR measures transferred feelings. The patient's parents lived in a different city; during the course of his therapy, the patient visited them twice, once between Hours 11 and 12 and once between Hours 26 and 27. Each visit home was followed by one or more hours in which he produced no ESS and then an hour in which ESSs are accompanied by an increased percentage of GSR. The first two-thirds of Hour 17 were devoted to sustained recollections of early home life in association to a dream, the only hour in the series in which such concentrated reminiscences occurred. The few ESSs in the last part of this hour were accompanied by an increased percentage of GSR. These facts suggest that the real or fantasied contacts with his parents restored in some degree the transferred feelings which had been partially extinguished during therapy.

Two interpretations of these results have not been sharply distinguished in this study. The feelings apparently measured by the GSR may simply be the fear aroused by acknowledging certain thoughts to oneself, rather than the fear connected with reporting them to another person such as the therapist. However, just as the fear of thinking these thoughts is probably acquired by generalization from fear of reporting them, so the extinction of fear during therapy would be expected to generalize from fear of reporting to fear of thinking. The two would probably covary, and for present purposes, it seems unnecessary to insist on the distinction. It is simply suggested that the fear of reporting, being the one which therapy primarily affects, provides the more likely interpretation of the results, but the alternative interpretation is not excluded by the design of this study.

SUMMARY

Statements by a therapy patient acknowledging personal sexual behavior or desires were frequently accompanied during early hours of therapy with galvanic skin responses, but not during later hours, following a progressive decline. This finding is taken as evidence that the GSR is associated with such feelings as fear or embarrassment in the interpersonal relationship with the therapist and that such feelings extinguish during therapy.

REFERENCES

1. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
2. FULTON, J. F. *Physiology of the nervous system*. New York: Oxford, 1943.
3. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
4. MOWRER, O. H., LIGHT, B. H., LURIA, A., & ZELENY, M. P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.) *Psychotherapy: theory and research*. New York: Ronald, 1953. Pp. 545-640.
5. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

Received April 27, 1956

THE EFFECTS OF THREAT AND FRUSTRATION ON GROUP COHESIVENESS

ALBERT PEPITONE AND ROBERT KLEINER¹

University of Pennsylvania

ACCORDING to recent field and laboratory studies, cohesiveness or "attraction-to-group" is a potent determiner of the strength of uniformity pressures (1), channels of face-to-face communication (3), the quantity of production (8), the expression of hostility (7), and other social behavior. There is, however, relatively little dependable information bearing on the factors that cause changes in cohesiveness. The present field experiment is concerned with this problem, namely, the effects upon cohesiveness of "threat" and "frustration."

Everyday observations of how these variables operate on cohesiveness are highly inconsistent. It is often apparent, for example, that groups under stress "pull together" and "close ranks" more than under normal conditions. On the other hand, there are circumstances such as battles where groups completely "fly apart." Beyond the fact that such observations are usually uncontrolled, they are difficult to interpret because of the multiple and sometimes dubious meanings attached at least implicitly to the concept of cohesiveness. Thus, changes in physical propinquity, uniformity of opinion, productivity, etc., which appear to result from threat and frustration do not necessarily mean a change in underlying cohesiveness—when the latter is strictly defined as attraction to the group. What adds to the difficulty is that the conditions of "threat" and "frustration" are defined in a variety of ways, and in many instances, appear to be regarded as conceptually and operationally equivalent. Finally, there seems to be an absence of even the most rudimentary theory with respect to this problem from which empirical predictions can be made.

Conceptually, "threat" is taken for present

¹ This study was made possible by a grant from the Albert M. Greenfield Center for Human Relations of the University of Pennsylvania. We also gratefully acknowledge the help of Mrs. Helen Leighty of the United Neighbors Association who made available the subjects and other facilities at Camp Linden, Northfield, Pa., where the study was carried out. Finally, the contributions of George Moed are appreciated.

purposes to refer to the probability that the group will sustain a loss in status. Different amounts of threat are thus represented by different probability statements with respect to a given degree of *loss*. The condition of "frustration," on the other hand, refers conceptually to the uncertainty over whether the group will *gain* a status position. Variations in the amount of frustration are thus also represented by different probability statements with respect to a given degree of *gain*. According to this conceptualization, the essential difference between threat and frustration is that the former is a motivational orientation exclusively or predominantly toward loss while the latter represents an orientation toward gain. One way of establishing these orientations is to vary the position of the group with respect to goals. Considering a hierarchical goal structure, groups that have already attained the highest possible goal position are potentially motivated against loss of their valued position. On the other hand, groups that have not yet attained such a position are potentially motivated toward gaining it.

The procedure implied by this reasoning was followed in the present experiment. Groups were divided into high and low status teams. The high status teams were then experimentally threatened by telling them of their chances of losing their position while the low status teams were experimentally frustrated by telling them of their chances of gaining the attractive position. Cohesiveness is defined in terms of interpersonal attraction among members of the team (1, 3, 4, 7, 8, 10). Operationally, the degree of cohesiveness is reflected by the number of positive sociometric choices made into the team by members of that team. Strictly the team by members of that team. Sociometric choices received by the team from other teams and choices given by the team to other teams.

As a general point of theoretical departure, the assumption may be made that cohesiveness is some function of the actual or potential "need satisfaction" that members obtain as a

result of their team experience (cf. 2, Ch. 7, for elaborations of this assumption). Next, team members are assumed to perceive themselves as responsible for providing each other with need satisfaction. Assuming, finally, that given expectations of gain and loss are equivalent to certain amounts of potential satisfaction of "approach and avoidance" needs, two hypotheses may be stated:

1. As the probability of loss decreases (i.e., as "threat" is reduced), cohesiveness increases.

2. As the probability of gain increases (i.e., as "frustration" is reduced) cohesiveness increases.

METHOD

The study was conducted during two successive sessions of a summer camp. The subjects in each camp session were boys seven to thirteen years of age, housed in eight cabin units. Since the experimental treatments and procedures were exactly the same in both camp sessions, the data were combined and thus represent the behavior of sixteen cabin groups. The assignment of the boys to each cabin was based upon similarity of age. In eleven cabin groups there were two Negro boys; in five cabins, the number of Negroes varied from one to four.

The Experimental Activities

On the first night of each camp session, the Program Director introduced the experimenters as coaches from the university who were at the camp to study the activity program. He then announced a tournament in which all cabins would participate. In substance, he gave the following description: The tournament will consist of three Preliminary Games followed by five Playoff Games. The games will include volley ball, touch football, and other familiar activities. Each cabin will be divided into two teams. These teams will compete against each other exclusively in the tournament games. The performance of each team will be scored on the basis of a multidimensional scoring system. The winners of the prelims will receive the title of Indian Chiefs and each Chief will be given an authentic Indian feather as a symbol of victory. The teams that lose the prelim games will be known as Warriors, and will not receive feathers. The Chiefs will enjoy various privileges at the end of the tournament; the Warriors will be denied various privileges. Finally, members of the teams that win the playoff games will each receive a pen-and-pencil set. Possession of the latter will be taken as a sign of consummate skill and prestige.²

Sociometric Measures

The sociometric measures were obtained from each cabin group separately on the second day of each camp session. Each group member was given a list of the names

of the boys in his cabin and the following instructions (paraphrased for brevity) were then given:

"The camp hopes that all of you in this cabin will be going on overnight hikes. They have asked us to find out the names of the boys in your cabin you'd like to go on hikes with, and those you wouldn't like to go with. We're going to ask you three questions. Before you answer, look over the list carefully and then tell us or write down the names of as many boys as you want for each question."

The questions were:

1. "If you were going on an overnight hike, who in your cabin would you most want to go with?" (The choices in this category are termed Positive choices.)

2. "If you were going on an overnight hike, and those you picked first couldn't go with you, who would you choose next?" (Lukewarm choices.)

3. "If you were going on an overnight hike, who in your cabin group would you not want to go with at all?" (Negative choices.)

Care was taken to identify each boy by name or nickname, and individual attention was given to insure that the instructions were clearly followed. The sociometric procedure was again administered just before the final playoff game. The instructions were exactly the same as the beginning sociometric administration, excepting for an introductory statement referring to the possibility of a final overnight hike.

Division of Groups Into Teams

Directly following the initial sociometric procedure, each cabin group was divided into two teams. The construction of the teams within the cabin groups and the assignment of the groups to the experimental treatments embodied several controls. First, the Negro members of each cabin group were distributed evenly between the two teams. It was possible to do this in thirteen of the sixteen groups. Second, the assignment of cabin groups to the experimental conditions was made so that the age ranges associated with the cabins were about equally represented under each experimental treatment. Unfortunately, we do not have data to show that in fact this equalization was approximately accomplished. Within the latitude permitted by these controls, an attempt was made to construct sociometrically equal teams within each of the groups and to assign the groups to the experimental treatments so that the latter would be sociometrically equivalent in terms of the number of positive and negative choices given by the team members to themselves. As will be seen later a perfect numerical equivalence was not achieved among the initial sociometric measures. Hence, in evaluating the difference in changes between experimental treatments, the fact that the changes involve different starting points has to be taken into account.

The Experimental Treatments

The treatments were administered directly after the establishment of the high and low status teams, that is, after the third prelim game.³ They were administered

³ To prevent suspicion about the scoring system the legitimate winners of the first Prelim Game became the High Status teams, i.e., won two of the three contests.

² All campers actually received this award at the end of the tournament.

in each cabin when both teams were present. In essence, the treatments were assertions by the experimenter directed simultaneously to the high and low status teams concerning their respective chances of losing (and winning) the tournament. The following excerpts should suffice to define these experimental treatments.

High Probability of Loss and Gain. "... but because they (*the high status team*) won doesn't mean they are finished and can keep the title of Indian Chiefs. If the Warriors (*the low status team*) win the playoffs, the Chiefs will have to surrender their titles and the tournament prizes. . . . We will now predict who will win in the end. We think that from the way things are going, the Warriors *will* win. . . . The Chiefs won the first two games but the point scores were very close. Today the Warriors won. So we predict the Chiefs will lose in the end."

It is essential to note that the same statement representing High Probability simultaneously functions to produce a strong expectation of loss in the high status teams and a correspondingly strong expectation of gain in the low status teams.

Low Probability of Loss and Gain. "... although the Chiefs are likely to win the playoffs, they shouldn't feel that it's all over. Don't forget the Warriors won today."

Observations

The behavior categories employed in the observation schedule were developed primarily on the basis of their theoretical relevance to the problem under study.

Insecurity concretely includes crying, expressions of fear and doubt, docility, and deference. Insecure behavior of this type is a common indicator of a state of threat. The category was employed, therefore, as an independent measure of the threat, i.e., probability of loss, variable.

Hostility concretely includes fist-fighting, spitting, marked aversion, and rage. Hostility has several theoretical linkages to cohesiveness which will be elaborated in the later discussion.

Rough Play concretely includes shoving, slapping, bumping, and similar items that lack hostile intent. Pilot observations indicate that rough play is a manifestation of competitive motivation. Common in close contests between rivals, it is behavior for which penalties are often meted out.

Group-Oriented Behavior concretely includes sharing of objects, giving assistance to each other, mutual planning, and other instances of cooperation and '*esprit*.' In view of the close connection between cohesiveness and cooperation, it is of interest to know whether this category of behavior will be affected by threat and frustration.

Self-Enhancement concretely includes bragging, fantasizing about power, showing off, etc. This category was included because observations of behavior in a city day camp suggest that it might be a consequence of frustration in play.

Power concretely includes giving commands, assign-

This is an admittedly uncontrolled factor in the experiment. However, the winners were "legitimate" primarily from *E*'s point of view. From the *Ss*' point of view, most of the games were very close contests.

ing roles, dictating "strategies," etc. The major basis for this category is the phenomenon of emergent leadership in stress situations.

Withdrawal concretely includes sitting down while the activity is going on, wandering off the field, etc. Similar behavior has been observed under frustration (5).

With the exception of the withdrawal category, the behaviors in the observation schedule can be directed toward members of the boy's own team or toward members of the opposing team. The interpretation of some categories obviously differs depending upon this directional factor.

The eight cabin groups in each camp session were observed during the prelim and playoff games the same number of times by each of the two observers. The average observation period was fifteen minutes, about one quarter of the time that the teams were engaged in each tournament activity. Each day each observer was assigned a different set of four cabin groups in a different temporal sequence.

During the observation period the observer recorded the precategorized behavior (symbolized by a letter) expressed to any one or more of the group members present. The initiator(s) and recipient(s) of behavior were recorded as well as the duration of the behavior in seconds.⁴ Since the team affiliation of both initiators and recipients was known it was possible to classify the behavior according to the team from which it emanated and to which team it was directed, i.e., whether it had an "own" or "other" direction.

The reliability of the observations was estimated on the basis of independent coding by the two observers of three different groups during the prelim and playoff games. Rank-order correlations of the frequency with which the two observers tallied the six categories were .80, .83 and .90. These coefficients, of course, only indirectly attest to the accuracy of the coding, i.e., the degree to which the same instance of behavior was placed in the same category.

The score used in connection with the observation data represents the difference in the frequency of behavior that a given team shows in a category between the prelim and playoff games. This difference is expressed as a percentage of the total behavior displayed by that team. Percentage scores were used, of course, to control for differences between teams in the absolute activity level.

RESULTS

Sociometric Measures

The average number of sociometric choices made by the high and low status teams at each sociometric administration are presented in Table 1. Here the choices are broken down according to whether they are positive, lukewarm, or negative; according to the experimental treatment under which they were made; and according to whether they are

⁴ Since some difficulty was encountered by the observers in keeping accurate time, the time indications are regarded as unit frequencies in the data analysis.

TABLE 1
AVERAGE NUMBER OF POSITIVE, LUKEWARM, AND NEGATIVE CHOICES

Direction and Time of Choice	High Probability			Low Probability		
	Pos.	Luke.	Neg.	Pos.	Luke.	Neg.
High Status Teams ($N = 8$)						
Own						
Initial		.18	.10	.32	.20	.09
Final	.39	.14	.13	.46	.13	.10
Differ.	.39		.03	.14*	-.07	.01
Other	.00*	-.04				
Initial		.16	.22	.38	.24	.17
Final	.51	.18	.16	.59	.23	.16
Differ.	.56	.02	-.06	.21	-.01	-.01
	.05					
Low Status Teams ($N = 8$)						
Own						
Initial		.16	.09	.30	.12	.16
Final	.39	.11	.12	.37	.17	.13
Differ.	.44		.03	.07	.05	-.03
Other	.05	-.05				
Initial		.23	.12	.37	.15	.21
Final	.51	.19	.16	.51	.12	.18
Differ.	.60		.04	.14	-.03	-.03
	.09	-.04				

* $F = 5.15$, $p < .05$, one-tailed test.

directed toward the teams themselves (own choices) or toward the opposing teams (other choices).

Because some teams did not have four members present at both sociometric administrations, a correction was necessary to make the own and other choices comparable from team to team. The correction consisted of dividing the number of choices made by the product of the number of team members who could make them and the number who could receive them.

A comparison of the amount of sociometric change produced by the experimental treatments can be made through analysis of covariance to adjust the differences between experimental treatments in final scores in proportion to the amount of regression of final upon initial scores. In effect, such a procedure compares the sociometric change across experimental treatment, i.e., the "net change," while making allowance for the unequal initial means.

The analysis of covariance was performed on the high and low status teams separately, since the effective, "psychological" quantities evoked by the high and low probability treatments are not necessarily equivalent for the high and low status teams. For example, high probability of loss is not necessarily equal in magnitude to high probability of gain. An additional theoretical reason for treating the

"loss" situation separately from the "gain" situation will be dealt with in a later discussion of results.

It will be recalled that the first hypothesis states generally that cohesiveness depends upon actual or potential need satisfaction that team members provide each other. Since low probability of loss presumably means relatively greater potential need (avoidance) satisfaction it can specifically be predicted that high status teams will show a greater increase (or lesser decrease) in own-team attraction under the low probability treatment than under the high probability treatment. The analysis of covariance confirms this prediction with respect to the positive choices. The F ratio representing the difference in final mean positive choices is 5.15 which is significant at less than the .05 level of confidence. None of the other differences in sociometric change for the high status teams attains statistical significance.

The prediction following from the second general hypothesis is that the low status teams under the high probability treatment will show a greater increase in own-team sociometric attraction than the low status teams under the low probability treatment. Table 1 shows that this prediction is not confirmed. Nor are any of the other comparisons statistically significant.

A serious weakness of the analysis just de-

TABLE 2
AVERAGE PROPORTION OF ACTUAL TO POSSIBLE SOCIO-
METRIC CHANGES IN PROCOHESIVE AND ANTI-
COHESIVE DIRECTIONS PRODUCED BY THE
HIGH AND LOW PROBABILITY
TREATMENTS

Direction of Choice	High Probability		Low Probability	
	Pro.	Anti.	Pro.	Anti.
High Status Teams (N = 8)				
Own	.42	.31*	.59	.12*
Other	.49**	.21	.67**	.14
Low Status Teams (N = 8)				
Own	.46	.18	.72	.29
Other	.56	.22	.53	.24

* $t = 2.33, p < .03$, one-tailed test.
** $t = 1.87, p < .09$, two-tailed test.

scribed is that it does not take *simultaneous* account of the changes in all three choice categories. It does not yield an over all measure of the change in cohesiveness brought about by the experimental treatments. To provide such a measure and one that controls for differences between treatments in the initial number of choices, the following procedure was adopted.

From a consideration of the choices or no-choices registered by each team member at the initial sociometric administration, the number of changes that theoretically could

occur in an upward ("procohesive") or downward ("anticohesive") direction was computed. Thus, a positive choice meant no possibility for upward change, but one possibility for a downward change. A lukewarm choice meant one possibility for a downward change and one possibility for an upward change. No choice meant one possibility for a downward (viz., to negative) and one possibility for an upward (viz., to either lukewarm or positive) change. The final sociometric data were then considered with respect to the changes that actually occurred to each initial choice or no choice. Finally, two proportions were computed for each team: (a) the ratio of actual to theoretically possible changes in a procohesive direction, and (b) the ratio of actual to theoretically possible changes in an anticohesive direction. Table 2 presents the averages of these scores under the various conditions.

Regarding the effect of the experimental treatments on the high status teams first, a significantly greater increase in anticohesive changes occurs under the high probability treatment. The low probability treatment, on the other hand, produced a larger but statistically insignificant increase in procohesive changes. Thus, this mode of analysis also confirms the first prediction: the low probability

TABLE 3
MEAN CHANGE IN PERCENTAGE OF BEHAVIOR IN SPECIFIED CATEGORIES TOWARD OWN AND OTHER TEAMS

Category	Status	High Probability (N = 8 groups)		Low Probability (N = 8 groups)	
		Own	Other	Own	Other
Insecurity	High	6.7*	-2.5	1.8*	6.9
Hostility	Low	-1.3	2.4	-1.5	6.2
	High	-2.7	-15.1**	3.8	12.2**
Rough Play	Low	-1.1	4.9	-3.1	1.1
	High	4.6	2.3	-4.5	-25.3
Group Oriented	Low	1.7	16.5***	-7.2	-11.1***
	High	2.8	3.7	1.8	1.1
Self-Enchantment	Low	2.2****	-10.0	18.9****	.2
	High	-10.2	5.6	-4.1	5.7
Power	Low	-5.1	-12.6	-11.9	-6.5
	High	-.7	4.8	3.9	-.5
Withdrawal	Low	2.7	-1.2	9.0	10.1
	High	-1.2	—	-2.4	—
	Low	-1.3	—	-1.5	—

Note.—The means superscribed by the same symbol are compared by t test with the following results.
* $t = 2.67, p < .02$, two-tailed test.
** $t = 2.60, p < .03$, two-tailed test.
*** $t = 2.25, p < .05$, two-tailed test.
**** $t = 2.29, p < .05$, two-tailed test.

treatment shows a lesser decrease in cohesiveness than the high probability treatment.

With respect to the changes in "other" sociometric choices, the *low status* teams under the low probability treatment became more attractive to the high status teams than they did under the high probability treatment. While this difference is not statistically significant, it is a trend that certain observation data appear to illuminate, and will therefore be discussed later.

Considering the low status teams, it is again clear that the second prediction is not confirmed. The high probability treatment does not produce a greater increase in cohesiveness than the low probability treatment.

Observation Measures

The observation data may now be examined for support and amplification of the findings thus far presented. Table 3 presents the relevant data on the mean change in relative frequency in each of the observation categories. The high status teams, first, show a significantly greater increase in insecure behavior directed toward themselves under the high probability treatment than under the low. They also show a significantly greater increase in hostility directed toward the low status teams under the low probability treatment than under the high. None of the remaining treatment comparisons of the high status teams is significantly different. With respect to the low status teams, Table 3 shows that a significantly greater reduction in rough play directed toward the high status teams occurred under the low probability treatment than under the high. The low status teams further show a significantly greater increase in group-oriented behavior directed toward themselves under the low probability treatment than under the high.

DISCUSSION

Concerning the effect of "threat" upon cohesiveness, the sociometric data clearly support the experimental prediction and allow us to state the general hypothesis that cohesiveness is an inverse function of the expectation of status loss. Confidence in both parts of this functional relation is strengthened when the observation data are examined. The fact that

the high status teams display more insecure behavior under the high probability treatment than under the low may be regarded as a direct reflection of the presumptive difference in the degree of threat aroused by the experimental treatments. In other words, this finding validates the independent variable. The greater increase in hostility directed toward the low status teams under the low probability treatment than under the high can, on the other hand, be interpreted in support of the dependent side of the hypothesis. It is well known that the amount of aggression released against an instigator depends upon the quantity of restraint operating against the group or individual in whom the hostility has been aroused. In a recent experiment, Pepitone and Reichling (7) have shown that cohesiveness enables groups to overcome such restraints and thereby facilitates the expression of hostility. The differential release of aggression, then, can plausibly be viewed as a "by-product" of, and hence the independent confirmation of, the predicted effects of the experimental treatment. It might be observed that the possible alternative argument that more hostility was *created* under the low probability treatment carries little force in view of the widely held supposition that hostility depends upon the amount of threat.

The greater release of hostility under the low probability treatment has an additional implication that might shed light on the marked (but not statistically significant) tendency for the high status teams under the low probability treatment to increase their "liking" of the low status teams more than under the high probability treatment (Table 2). Two experiments suggest the reason for this occurrence. Thibaut (10) has shown that teams which move into high status positions and express hostility freely toward the teams they have displaced do not show the expected reduction in the number of positive sociometric choices made toward the displaced teams. The overt expression of hostility apparently dispels the negative feelings that otherwise would have resulted in fewer positive sociometric choices. Pepitone and Reichling (7) have shown that groups which released a large amount of hostility toward an instigator afterward regarded him more favorably than groups which

released relatively little aggression. Thus, available evidence plausibly suggests that a "catharsis" of hostility was responsible for the greater attractiveness of the low status teams (for the high status teams) under the low probability treatment than under the high.

The generality of the hypothesis concerning the effect of "threat" upon cohesiveness warrants consideration. It was earlier assumed that for cohesiveness to be affected by reduced expectation of loss, team members must perceive each other as responsible for the reduction in status. Such an assumption implies that the increase in own-team attraction represents expressions of appreciation and gratitude on the part of the team members toward each other for having reduced the danger that threatened them, and that changes in cohesiveness may not result when the reduction in threat is perceived to originate outside the team. Such a "causal attribution" could well be the reason why the high and low threatened groups in the experiment by Schachter *et al.* (9) did not differ in cohesiveness as measured by the question: "How frequently do you think this group should meet?" In this experiment, the threat originated in the announcement that a number of groups would have to be eliminated from a contest because there were insufficient resources. The threat was wholly external to the group members and they had no responsibility for it.

Turning now to a consideration of the low status teams, the data of the present experiment do not confirm the prediction made with respect to the consequences of differential "frustration." The possibility that the *objective* difference in probability of gain did not create *psychologically* different treatments for the low status teams seems remote in view of the systematic differences in observed behavior already noted. There are, however, two theoretical possibilities that can be considered seriously.

It is possible that low status teams did not tend to see themselves responsible for changes in the expectation of gain, and the differential expectation of gain between the experimental treatments was therefore not reflected in the sociometric choices. There is evidence in support of this interpretation. Thibaut and Riecken (11) found that low status Ss tended to perceive themselves as relatively irresponsible for the overt compliance they achieved in the

high status confederates. To generalize this finding, low status groups tend to see changes in their own situation as being due relatively less to their own performance and relatively more to external sources such as the Experimenter, "luck," opposing groups, etc.

A second theoretical possibility to account for the no-difference in cohesiveness between the high and low probability treatments requires discussion of the observation finding relating to the low status teams. The greater amount of "other-directed" rough play under the low probability treatment can be straightforwardly interpreted. According to the definition of the rough play category given earlier, this finding indicates that when the probability of gain is relatively low there is a tendency for competitive activity to be reduced. The greater amount of "own-directed" group-oriented behavior may be plausibly viewed as coordinate with the finding on rough play. Thus, the lower the probability of gaining a status position, the greater the tendency for team members to withdraw from competition with the other team *and* to enter into cooperative relations with each other.

Perhaps these interpretations can be rendered more familiar when the experimental treatments, high and low probability of gain, are considered as degrees of barrier strength. It is commonly observed that when barriers are very strong there is a tendency to cease goal striving. Level-of-aspiration studies show that such "giving-up" occurs specifically when the probability of success precludes the attainment of the stated goal level (6). Data from the present experiment appear to show this reaction in the case of a group being frustrated.

We may now return to the question of why there was no difference in cohesiveness between the high and low probability-of-gain treatments. A purely speculative explanation is that the withdrawal from competition and the increased cooperation on the part of the low status teams under the low probability treatment were themselves need-satisfying behaviors in that they reduced the frustration. While initially the high and low probability treatments differed in potential need satisfaction, the actual need satisfaction eventually obtained by the low probability teams cancelled this difference. Hence, the treatments did not produce differential cohesiveness.

SUMMARY

A field experiment was concerned with the effects upon cohesiveness of "threat" and "frustration," the latter more precisely defined in terms of expectations of loss and gain in status respectively. On the basis of sociometric and behavior observation data, the following interpretations seem warranted:

1. A reduction in the expectation of status loss (a reduction in threat) tends to increase cohesiveness.

2. An increase in the expectation of gain (a reduction in frustration) does not produce an increase in cohesiveness.

3. There is some evidence that a catharsis of hostility increases the attractiveness of the teams toward which the hostility is directed.

4. There is evidence that low expectation of gain (strong frustration) produces a withdrawal symptom consisting of reduced interteam competition and increased intrateam cooperation.

REFERENCES

1. BACK, K. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, 46, 9-23.

2. CARTWRIGHT, D., & ZANDER, A. (Eds.), *Group dynamics*. Evanston, Ill.: Row Peterson, 1953.
3. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups*. New York: Harper, 1950.
4. FESTINGER, L. Architecture and group membership. *J. soc. Issues*, 1951, 7, 152-163.
5. FRENCH, J. R. P. Organized and unorganized groups under fear and frustration. *Univer. of Iowa Stud. Child Welf.*, 1944, 20, 231-307.
6. LEWIN, K. et al. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 333-378.
7. PEPITONE, A., & REICHLING, G. Group cohesiveness and the expression of hostility. *Hum. Relat.*, 1955, 3, 327-339.
8. SCHACHTER, S., ELLERTSON, N., MCBRIDE, D., & GREGORY, D. An experimental study of cohesiveness and productivity. *Hum. Relat.*, 1951, 4, 229-238.
9. SCHACHTER, S., NUTTIN, J., DE MONCHAUX, CECILY, MAUCORPS, P. H., OSMER, D., DUIJKER, H., ROMMETVEIT, R., & ISRAEL, J. Cross-cultural experiments on threat and rejection. *Hum. Relat.*, 1954, 4, 403-440.
10. THIBAUT, J. W. An experimental study of the cohesiveness of underprivileged groups. *Hum. Relat.*, 1950, 3, 251-278.
11. THIBAUT, J. W., & RIECKEN, H. W. Some determinants and consequences of the perception of social causality. *J. Pers.*, 1955, 2, 113-133.

Received May 4, 1956

EFFECTS OF PERSONAL AND IMPERSONAL REFUTATION OF AUDIENCE COUNTERARGUMENTS ON ATTITUDE CHANGE

JOSEPH KAMENETZKY AND HANS SCHMIDT¹

University of Illinois²

PERSUASIVE communications frequently contain refutations of opposed arguments and points of view. A question that arises in the preparation of such communications concerns whether it is more effective to express refutations as if they are directed primarily at the opposed members of the audience (personal refutation) or as if they are aimed at some abstract point of view (impersonal refutation). The latter, less direct, strategy might be expected to be superior if refutative appeals formulated as disproofs of the audience's beliefs seem likely to antagonize the audience. Contrariwise, the former strategy should tend to be more effective insofar as opposed members of the audience are less likely to grasp a refutation that is indirect or subtle (3, 4, 7). These expectations may be formulated in terms of the following hypotheses:

1. Other things being equal, if personal refutation antagonizes opposed members of the audience more than impersonal refutation, personal refutation is less effective in modifying opinion.

2. Other things being equal, if personal refutation is more informatively adequate among opposed members of the audience than impersonal refutation, personal refutation is more effective in modifying opinion.

METHOD

Subjects. Subjects (Ss) were 216 college men enrolled in the introductory speech course at the University of Illinois during the spring semester of 1955. The Ss reported to a special experimental classroom in groups of approximately 13. Two aggregates of five groups and one aggregate of six groups were assigned randomly to

¹ The writers are grateful to Professor Donald L. Thistlethwaite for his generous assistance. They also wish to thank the faculty and students of the Department of Speech, University of Illinois, for their wholehearted cooperation.

² This research was supported in part by the United States Air Force Contract No. AF 33(038)-25726, monitored by the Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States government.

one of the two experimental treatments or to the control treatment.

Communications. It was desirable that a controversial topic be selected which satisfied the following criteria: (a) the majority of the audience should be opposed initially to the communicator's proposition; (b) the most salient audience counterarguments to this proposition should be known. A survey conducted during the previous semester suggested that most introductory speech students were opposed to abolishing the program of student deferments and that four more or less distinct and salient arguments supporting their opposition could be ascertained.

Two scripts were prepared, each of which recommended that the Selective Service System *abolish* all student deferments. Both scripts contained the same facts and arguments in support of this viewpoint. Audience counterarguments were neither elaborated nor explicitly acknowledged. Instead, each explicit refutation was followed by a presentation of supporting facts and the refutation was then restated. The four refutation sequences asserted that abolishment of student deferments would not (a) adversely affect the military services, (b) discourage men from completing their education, (c) be harmful to the student's subsequent academic performance, or (d) endanger the nation's supply of scientific and technical manpower. All refutations were summarized in the final paragraphs of the communications.

The personal and impersonal refutation scripts differed primarily in that the second person pronoun *you* was substituted 27 times for third person nouns or pronouns. A typical sequence in the impersonal refutation script was as follows: "How many people can point to a single skill which they will attain in college that could be put to immediate use in military service? Very few, if any, could." The corresponding sequence in the personal refutation script was: "How many of *you* will have attained in college a single skill that could be put to immediate use in military service? Very few of *you*, if any, could."

Transcriptions were made by a student announcer for the University of Illinois radio station who was introduced as a former consultant for the National Manpower Commission. Running time of each recording was approximately 18 minutes.

Control Ss listened to an unrelated communication on factors in the rise of anti-intellectualism in America.

Questionnaire materials. At the start of the experimental session each S ranked the following sentences in order of preference, i.e., in the order in which the sentences corresponded to his own personal position:

1. All men in college should be deferred from the draft.
2. The majority of college students should be given deferments.

3. No more than *half* of the men in college should be deferred.
4. Deferments should be given only *rarely* and only under exceptional circumstances.
5. All student deferments should be abolished.

When *pretest* preferences were scaled by the unfolding technique (2), only 87 per cent of the observed rank orders were found to be deducible from the best fitting joint scale. However, when only the first *two* ranks were considered, the unfolding analysis classified 98 per cent of the rank orders on the joint scale (cf. 2, p. 515). Accordingly, the latter scale, which provided eight distinguishable levels of favorableness toward abolition of student deferments, was used in the subsequent analyses.^{3, 4}

Immediately after exposure to the talk, a 15-item questionnaire was administered. Ten items dealt with opinions on the adequacy, relevance, and truthfulness of the speaker's factual presentation, his freedom from bias, whether *S* believed the talk should be given to other *Ss*, etc. Responses to these items are hereafter called *discounting* reactions. A modification of Guttman's scaling technique described elsewhere (6, 7) was used in scaling the discounting responses. All ten questionnaire items were included in, and reduced to, a five-item contrived scale with a coefficient of reproducibility of .93. Each item in the contrived scale comprised one, two, or three questionnaire items.

Five multiple-choice items designed to measure the extent to which *S* could recognize the intended conclusion of the speaker were also included. The number of correct responses gave an indication of *S's* *comprehension* of the speaker's major conclusion.

The final questionnaire presented 14 key attitude items concerned with the abolition of student deferments. When responses to these *posttest* items were scaled by the same method, a six-item scale with a reproducibility coefficient of .94 was obtained. In the following analyses a high score indicates a favorable attitude toward abolishing student deferments.

The *Ss* completed all questionnaires anonymously. For exploratory purposes, they also registered their affective responses to the communications on a "program analyzer" equipped with three-way switches. (These data were not used in the analysis to be reported.) Switch numbers, instead of names, were used in collating the questionnaires.

³ The product-moment correlation between *pretest* preferences and *posttest* scores for control *Ss* was .74. Rankings to an additional set of five statements were also obtained. However, fewer *Ss* could be classified on the Joint scale and the correlation of these scale scores with *posttest* scores was only .65.

⁴ Although each of the predictions concerns effects upon opposed *Ss*, no distinction between those initially favorable and initially opposed was made in the analysis. Since only 10 per cent of the sample had scores falling in the lower four of the eight levels of favorableness toward abolishment, it was assumed that the *Ss* were almost unanimously opposed to the speaker's recommendation.

RESULTS AND DISCUSSION

Since the treatment groups differed somewhat in distributions of initial attitude levels and of initial rate of "dislike" affective responses, each treatment group was matched on these variables. Whenever more than one individual could be matched with another, the selection was made by referring to a table of random numbers.

Effects on arousal of antagonism and comprehension and on attitude change. To take into account the possibility that some variance in the discounting and comprehension scores can be attributed to initial attitude, a treatments $\times$ levels analysis of variance of the discounting and of comprehension scores was employed. Each *pretest* distribution was divided into four levels of 9, 15, 18, and 7 *Ss* per level. In neither case was there a significant difference between the refutation groups in over-all appraisal of the communications or in comprehension of the conclusion.

Insofar as the personal and impersonal refutation treatments did not differ significantly in the degree of antagonism elicited or in comprehension of the intended conclusion, no prediction concerning the relative effectiveness of the refutation manipulations upon attitude change was possible in terms of the hypotheses of the study. The results of treatments $\times$ levels analysis of variance of *posttest* attitude scores of all three treatments indicated that the major source of variance of the *posttest* attitude scores was initial attitude level; the levels difference was significant at the .001 level. The *F* value corresponding to the mean square for treatment was significant at the .05 level.

Linear comparisons between selected treatment groups were made by a method described by Cochran and Cox (1, pp. 56-59). The results showed that the means of each of the refutation treatments, as well as that of the combined experimental treatments, differed from the control group mean at the .05 level. The experimental communications were apparently effective in increasing favorableness toward abolishing student deferments. The difference between personal and impersonal refutation was not significant. Therefore, it cannot be concluded that either method of refutation was more effective.

TABLE 1
ADJUSTED POSTTEST MEASURES OF ATTITUDE TOWARD
THE ABOLITION OF STUDENT DEFERMENTS FOR
SUBJECTS WITH HIGH AND LOW DISCOUNTING
SCORES

Treatment	Ss with High Discounting Scores			Ss with Low Discounting Scores		
	N	Error of estimate	Adjusted mean	N	Error of estimate	Adjusted mean
Personal refutation	22	1.01	2.30	27	.98	3.44
Impersonal refutation	27	1.62	2.89	22	2.28	2.92
p diff.*			.04			.04**

* One-tailed test.

** This probability value should be interpreted with caution since the assumption of homogeneity of variance is not fulfilled ($p < .05$). The means are listed primarily to illustrate the meaning of the partial correlation.

Correlations between dependent response measures. The two hypotheses imply that attitude change is (a) inversely related to the extent to which the communication antagonizes the audience, and (b) positively related to comprehension of the speaker's conclusion. While no differences between the refutation treatments with respect to arousal of antagonism, comprehension, or attitude change were demonstrated, relationships between attitude change and the other two variables can still be investigated correlationally. Partial correlations were therefore computed between posttest attitude and comprehension scores, holding pretest attitude constant. The partial correlation was +0.04 for the personal refutation treatment and +0.22 for the impersonal refutation treatment; for both treatments combined the correlation was +0.14. None of these correlations is significant at the .05 level. Comprehension as measured here does not appear to be a *sufficient* condition for attitude change. Partial correlations were also calculated between posttest attitude and discounting scores, holding pretest attitude constant. The correlations for the personal and impersonal refutation groups were -0.44 and +0.09, respectively. The former correlation is significantly different from zero ($p < .01$); the difference between the two correlations is also significant ($p < .01$). Hence it may be concluded that only under certain conditions are discounting tendencies related to attitude change.

To shed light on the meaning of the differ-

ences between the two correlations, Ss were classified as to whether they scored above or below the group median on the scale for measuring discounting tendencies. For the high and low discounting subgroups, respectively, the mean posttest attitude scores for the personal and impersonal refutation treatments were compared by covariance analysis with level of pretest attitude controlled. The results (Table 1) show that the high discounters in the personal refutation treatment had a significantly *lower* mean posttest attitude score than the high discounters in the impersonal refutation treatment, while the low discounters in the personal refutation treatment had a significantly *higher* mean posttest score than the low discounters in the other treatment. In both comparisons variability was less for the personal refutation treatment although significantly so ($p < .05$) only among the low discounters.

One testable interpretation of this finding is as follows. *Personal* refutation has the effect of leading recipients to believe that the refuted arguments are the basis of their opposition to the speaker's conclusion. Those (low discounters) who accept the adequacy of the speaker's arguments and refutations will tend to exhibit positive attitude change, while those (high discounters) who challenge the cogency of these same arguments and refutations will tend to show little or no attitude change. Recipients of the *impersonal* refutation, on the other hand, are relatively less prone to recognize the refuted arguments as their own. Hence they may feel less obliged to be "consistent" in their reactions. This supposition can be readily tested by having Ss in both treatments evaluate, after hearing the communication, the relevance of the speaker's assertions to their own position.

In the present experiment, personal refutation produced no greater attitude change or discounting tendencies than did impersonal refutation. The correlation between these two dependent variables for personal, but not for impersonal, refutation suggests that personal refutation would be the more effective mode if the over-all level of discounting reactions were reduced. This reduction might be accomplished by employing an additional variable such as prestige (cf. 5) or by adequately refuting more

salient audience counterarguments.⁵ Under these conditions we would expect a differential increase in attitude change for the personal refutation over that for impersonal refutation. Similarly, the use of a variable which serves to increase discounting reactions would be expected to result in greater effectiveness of impersonal refutation.

SUMMARY

The present experiment compared the effectiveness of personal and impersonal modes of refuting audience counterarguments in changing attitudes toward the abolition of student deferments. There was no evidence that personal refutation was any more or any less effective in this respect than impersonal refutation; nor were there any significant differences in the extent to which recipients (a) comprehended the speaker's intended conclusion or (b) discounted the talk as being biased or unfair. For the personal refutation group the more a subject discounted the appeal, the less favorable he was to the speaker's

proposition; this relationship did not obtain for the impersonal refutation group. The experimental implications of this finding are discussed.

REFERENCES

1. COCHRAN, W. G., & COX, GERTRUDE M. *Experimental designs*. New York: Wiley, 1950.
2. COOMBS, C. H. Theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953. Pp. 471-535.
3. COOPER, EUNICE, & JAHODA, MARIE. The evasion of propaganda: How prejudiced people respond to anti-prejudice propaganda. *J. Psychol.*, 1947, **23**, 15-25.
4. KENDALL, PATRICIA L., & WOLF, KATHERINE M. The analysis of deviant cases in communications research. In P. F. Lazarsfeld & F. N. Stanton (Eds.), *Communications research 1948-1949*. New York: Harper, 1949. Pp. 152-179.
5. OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42-55.
6. STOFFER, S. A., BORGATTA, E. F., HAYS, D. G., & HENRY, A. F. A technique for improving cumulative scales. *Publ. Opin. Quart.*, 1952, **16**, 273-291.
7. THISTLETHWAITE, D. L., & KAMENETZKY, J. Attitude change through refutation and elaboration of audience counterarguments. *J. abnorm. soc. Psychol.*, 1955, **51**, 3-12.

Received May 4, 1956.

⁵ Actually, the saliency of the four counterarguments refuted in this experiment was only moderate. No counterargument was given by more than 20 per cent of Ss questioned in the preliminary survey.

AUTONOMY AND JOB SATISFACTION IN TASK-ORIENTED GROUPS¹

DONALD B. TROW

Carnegie Institute of Technology

Two of the fairly consistent findings in studies of laboratory groups with systematically varied communication nets (i.e., patterns of communication links or channels between pairs of members) are that the job satisfaction reported by a group member and the status attributed to him by other members vary directly with the centrality of his position in the net (2, 4, 5, 7). Centrality, roughly speaking, is the degree to which a person's position in an organization is unaffected by the requirement that communication go through "channels." Its basic measure, suggested originally by Bavelas (1), is the sum of distances (i.e., minimum numbers of communication links) from *each* position to every other position divided by the sum of distances from the *given* position to each of the others. The purpose of the present experiment was to test the possibility that the empirical relationship of centrality to satisfaction and status may be accounted for in large part by two variables: (a) the autonomy of positions in the organization and (b) the members' need for autonomy.

Autonomy is here defined as the degree to which a person's position in the information-flow of an organization permits him to determine for himself the organizationally appropriate level or direction of his own future activity. The concept designates the degree of independence from others with respect to direction, not to information; a person occupying a position of high autonomy may be—and in fact usually is—highly dependent upon others for information about their activities, the state of the environment, etc.

Although autonomy is conceptually independent of centrality, the one referring to access to task-relevant information and the other to access to information channels, an actual sample of task-oriented groups and organizations would probably show them to

be positively correlated.² But because not all information is task-relevant and not all task-relevant information is communicated, instances of a high degree of one with a low degree of the other do occur. A company switchboard operator, for example, presumably has a high degree of access to organization channels but must continually be told which connections to make.

Autonomy is, however, closely related to another concept, that of "answer-getting potential" used by Leavitt to explain his obtained relationship between centrality and satisfaction among groups doing the common-symbol problem (5). Answer-getting potential, to judge from Leavitt's discussion, might be defined as the degree to which a person's position in the information-flow of an organization permits him to determine for himself which piece of information (symbol) is held in common by all group members, and thus differs from autonomy only in the kind of task to which the information is relevant.

The present experiment attempts to test Leavitt's explanation of his findings, and at the same time to generalize it beyond the answer-getting type of task. Leavitt's explanation ran essentially as follows: If members initially send messages approximately equidistantly to all potential recipients (excluding those who are known to have all the information the message might contain), then, by chance, centrally located members are more likely than peripheral members to be the first to receive all the items of information and thus to obtain the answer; central members then can and do regularly send out the answer to their less fortunate peripheral brethren; this irregularity becomes perceived and, in conjunction with cultural norms that place high value on answer-getting, affords high status to central members; finally, the perception by a central member of his higher status and/or

¹ This article is based on the author's Ph.D. dissertation completed at the University of Michigan in 1955 under the chairmanship of T. M. Newcomb, whose help is gratefully acknowledged. Parts of this paper were read at the American Psychological Association meetings in Chicago in 1956.

² The term "task-oriented" is used to distinguish such groups from purely social groups—although the term can be taken in the more precise sense of Christie, Luce, and Macy (2)—and does not necessarily imply task- rather than ego-involvement on the part of individual members.

the direct perception of greater self-sufficiency yields a high level of job satisfaction.

This explanation could not be tested in Leavitt's data because the postulated answer-getting potential was perfectly correlated with centrality. In the present experiment the test is accomplished by varying these independently, at the same time measuring, in terms of individuals' needs for autonomy, variation among individuals in the force of the alleged cultural norm. In addition the task is changed to one in which the deduction from accumulated pieces of information, instead of being merely an "answer," is a new optimal direction for the member's on-going activity. It is for this reason that the term "autonomy" has been substituted for "answer-getting potential."

The specific hypotheses of this study, then, are as follows:

1. The greater the autonomy of a person's position, the greater his job satisfaction tends to be;
2. The stronger a person's need for autonomy, the more highly is his job satisfaction correlated with the autonomy of his position; and
3. The greater the autonomy of a person's position, the higher the self-ascribed status of the position.

METHOD

Subjects

Forty-four volunteer college students served as subjects (Ss). Three were scheduled for each experimental session which lasted about one hour. If only two appeared, an assistant filled in as the third S, since it was necessary that the groups appear to be of size three but not necessary that all three be bona fide Ss. A routine check was made to be sure the three were not acquainted.

Procedure

Each S filled in a need questionnaire (see below), and as soon as all were finished E read the instructions, which made the following points: (a) Each member of the group would be assigned a series of tasks—outlines of animals to be cut out of paper; (b) a person might be given a task identical with that of another member; (c) the goal for the group would be to finish as many different tasks as possible in as short a time as possible, thus making it a waste of time for a person to do a task identical to one which another member had already done or was doing; (d) the members would have to communicate with each other via written notes delivered by E and refer to the animals only by the nonsense names they would be given; (e) identical tasks might

be given different names, but (f) one member would be given the "translations" so that this information would be available to help prevent duplication of effort. To produce the effect of a series of tasks—and thus a continuing need for coordinative communication—Ss were instructed to cut out the patterns in the order in which they are initially arranged (although they might skip any which they knew another member had done or was doing) and restricted in sending out task information to communicating only the (nonsense) name of the animal they were currently cutting and the next one in order. The Ss were further told, in effect, that although the tasks would be simple and the restrictions on communication might seem artificial, the experimental conditions constituted a laboratory model of situations that could be found in all organizations. As soon as any questions of clarification about the instructions had been answered, Ss were shown to separate cubicles and given their working materials, which included a card showing each his assigned letter and the letter of the person to whom he was allowed to write notes ("W" and "X," respectively, for all Ss). At a common signal each S began work on the first of the six tasks or on writing a first note, as he chose. Through notes ostensibly from other members but actually prewritten by the experimenter, each S was led to believe that he (W) and another member (Y) had two tasks in common and that the third member (X) had the "translations." Actually, all Ss had the same series of tasks and—in any one variation—received identical notes. The time of delivery of notes was coordinated with each S's progress in cutting out his series of animals, the occasional visits to see if he had any outgoing notes providing sufficient opportunity to see whether he was far enough along for delivery of a note on the next trip. When S signalled that he was finished, his time was recorded and he filled in a postexperimental questionnaire. When the last one had finished his questionnaire, the session was over and its purposes and methods were fully explained.

Measurement of Needs

In the pre-experimental need questionnaire S was shown five statements of social needs (Achievement, Autonomy, Affiliation, Recognition, and Cognition) introduced as "guidelines or principles" that might be followed in one's day-to-day life, and was asked to rank and rate them according to their importance to him.³ Rankings and ratings of the Autonomy statement ("Carrying out the business of life in an independent and self-reliant fashion") were combined to divide the Ss, for analysis, into two equal groups having relatively high and relatively low need for autonomy. Analysis of data from 33 pretest Ss had shown the rankings and ratings to be relatively stable over a two-months' interval (median τ for the rankings = .6; test-retest r for Autonomy ratings = .73) and independent of the order of presentation of the statements.

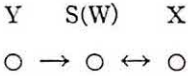
Experimental Variations

Four experimental variations were introduced in the content of the prepared notes S received. He was

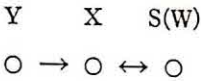
³ This questionnaire was devised by A. R. Cohen (3), and permission to use it is gratefully acknowledged.

led to believe he occupied either a central or peripheral position in the communication net and was made either autonomous or dependent upon "X" for decisions as to which animals not to cut out in order to avoid duplication. The operations used to produce the perceptions of centrality or peripherality were as follows:

Central variations. S receives notes ostensibly from X and from Y. In addition, X's first note states: "I can write only to you," and Y's second and third notes convey the information that he is not receiving notes from anyone. This information, together with S's instructions that he can write only to X, defines the following net:



Peripheral variations. S receives notes ostensibly from X only. X's first note again states, "I can write only to you," and his second and later notes relay information from Y, indicating that Y can write only to him. This information and S's instructions define the net:



In all variations S receives and sends notes from time to time of the type, "I am now cutting out a ____." However, the notes received contain no useful information in themselves because the nonsense name in the blank is never the name of any of the animal figures S has, although there is a possibility, emphasized in the instructions, that two identical figures may have different names. The instructions specifically forbid referring to an animal in any way except by its nonsense name, so that the promised translations become crucially important to S's decision with respect to each figure.

Autonomous variations. The first note received from X says in part, "Here is the translation" and proceeds to list equivalences of pairs of nonsense syllables, some of which are names of animals S has. The translations plus the incoming notes telling what figures others are cutting give S all the information necessary to decide whether or not to cut out each figure as he comes to it. The notes are timed in such a way that in the middle of cutting out the third and fifth animals he learns that Y at some earlier time was cutting out the same animal—though under a different name.

Dependent variations. The first note from X, instead of containing the translations, informs S only that X has them and (to discourage requests for them) that "It's a long list." Thus when S hears which figures (untranslated) others are doing, he must ask and/or be told by X whether these are the same as any of his own and thus whether or not he should skip any.

In all variations S receives two notes from X saying, "Don't cut out the ____." In the dependent variation this is in reply to S's question if he asks it, otherwise gratuitous. In the autonomous variation it is superfluous direction, since the note is delivered only after S has shown he is not going to cut out the animal in question.⁴

⁴ Eight of the Ss in the autonomous variations fin-

TABLE 1
JOB SATISFACTION IN THE FOUR EXPERIMENTAL VARIATIONS
(N = 11)

Variation	Mean Satisfaction*	SD
Central-Autonomous	1.18	.39
Peripheral-Autonomous	1.91	1.31
Central-Dependent	2.18	.83
Peripheral-Dependent	2.27	1.14

Analysis of Variance			
Source	df	Variance Est.	p
Autonomy-Dependence	1	5.11	<.05
Centrality-Peripherality	1	1.84	
Interaction	1	1.11	
Error	40	1.06	

* 1 = liked it very much; 2 = liked it fairly well; 3 = liked some things about it, disliked others; 4 = didn't like it too well; 5 = didn't like it at all.

The combination of the two dichotomous variables produces four experimental variations: central-autonomous, central-dependent, peripheral-autonomous, and peripheral-dependent. The number of notes, their sequence, and—with the exceptions noted above—their content and time of delivery were the same in all variations.

Answers to questions on the postexperimental questionnaire indicate that the induction of perceived centrality or peripherality was successful for all Ss. Autonomy was apparently less readily perceived, for a substantial number of Ss (14 out of 22) in the autonomous variations checked that they had been "following directions as to which patterns to cut," rather than "deciding for myself." However, some or all of the 14 may have interpreted "directions" to include those from E, who had in fact given directions as to which animals to cut—namely, those not being cut by someone else.

RESULTS

The first question on the postexperimental questionnaire asked: "Considering it as a whole, how much did you like your job?" Table 1 shows the mean satisfaction (liking) in each of the experimental conditions and the results of analysis of variance. It can be seen that situational autonomy produces a sig-

nificant effect. Eight of the Ss in the autonomous variations finished cutting out all animals even though they possessed sufficient information to indicate that the two of them were being (or already had been) cut out by Person Y. Several of these Ss indicated afterwards that they had not checked incoming notes against the "translations" and so had not realized they were duplicating Y's work. Data from these eight were discarded and sufficient additional Ss run to equalize cell totals.

TABLE 2
JOB SATISFACTION IN AUTONOMOUS AND DEPENDENT
VARIATIONS BY LEVEL OF NEED
FOR AUTONOMY

n Autonomy	Experimental Treatment						t	p
	Autonomous			Dependent				
	n	Mean	SD	n	Mean	SD		
High Low	10	1.60	.80	12	2.50	.96	2.32	.05
	12	1.50	1.19	10	1.90	.94	.26	n.s.

TABLE 3
RELATIVE STATUS OF X IN THE FOUR EXPERIMENTAL
VARIATIONS
(N = 11)

Variation	Mean Status*	SD
Central-Autonomous	2.64	.88
Peripheral-Autonomous	1.91	.90
Central-Dependent	2.27	.75
Peripheral-Dependent	1.45	.50

Analysis of Variance

Source	df	Variance Est.	p
Autonomy-Dependence	1	1.84	.01
Centrality-Peripherality	1	6.57	
Interaction	1	.023	
Error	40	.66	

* Mean status of X relative to self: 1 = much higher; 2 = somewhat higher; 3 = same; 4 = somewhat lower; 5 = much lower.

nificantly higher level of job satisfaction than does situational dependence, as predicted by Hypothesis 1, and that the effect of centrality, though positive, is not significant. The effect of autonomy on job satisfaction was no greater among those who checked "deciding for myself" on the postexperimental questionnaire than among those who checked "following directions."

For testing of Hypothesis 2, the Ss were divided into equal groups having high and low need for autonomy. Table 2 shows mean satisfaction of Ss high and low in need-level in autonomous and dependent situations. For high-need Ss the difference in mean satisfaction as between the two situations is significant at better than the .05 level, while for low-need Ss it is not significant, a result in accord with the hypothesis.

The relationship posited in Hypothesis 3 between autonomy and status was exactly comparable to that between autonomy and

satisfaction. One of the last questionnaire items asked, "If this were a commercial enterprise, how much do you think X should be paid in comparison with the person filling the position you had?" Table 3 shows the mean response to this item in each of the variations, and the results of analysis of variance. There are two things which may be noted about these findings. First, the average S in all variations sees X as deserving higher pay than himself, even in the variation where S is both central and autonomous. This finding is quite probably the result of X's sole possession of the translations initially, although it was intended that, in the autonomous variations, their initial location would be perceived as a matter of little importance. Secondly, when X is central and S peripheral, the pay differential judged to be fair by S is particularly large, while differences in autonomy seem to have no effect. In contrast to job satisfaction, status judgments appear to depend upon centrality-peripherality rather than upon autonomy-dependence, and Hypothesis 3 fails of confirmation.

Despite the lack of effect of situational

TABLE 4
PERCEIVED STATUS AS A FUNCTION OF INTERACTION
BETWEEN NEED FOR AUTONOMY AND
CENTRALITY-PERIPHERALITY
(N = 11)

(IV - II)						
n Autonomy	Experimental Treatment				t	p
	Central		Peripheral			
	Mean	SD	Mean	SD		
High	2.55	.89	1.91	.90	1.60	.20
Low	2.36	.77	1.45	.50	3.14	.01

TABLE 5
PERCEIVED STATUS AS A FUNCTION OF INTERACTION
BETWEEN NEED FOR AUTONOMY AND
SITUATIONAL AUTONOMY-DEPENDENCE

SITUATIONAL AUTONOMY-DEPENDENCY								
n Autonomy	Experimental Treatment						t	p
	Autonomous			Dependent				
	n	Mean	SD	n	Mean	SD		
High	10	2.88	1.05	12	2.27	.62	1.61	.20
Low	12	2.46	.64	10	2.25	.65	.73	n.s.

* Adjusted to control for centrality-peripherality by adding to the scores of Ss in peripheral variations the mean difference between central and peripheral variations—for the high and low need groups separately.

autonomy, need for autonomy appears to influence status judgments. As shown in Table 4, it is only among the low-need Ss that differences between central and peripheral variations in status judgments reaches a high level of significance. Those Ss with a strong need appear reluctant to ascribe status on the basis of centrality-peripherality; moreover, there is a tendency, which does not reach an acceptable level of statistical significance, for their status judgments to be influenced by situational autonomy-dependence instead (Table 5).

CONCLUSIONS AND DISCUSSION

The following conclusions may be drawn from this experiment:

1. The job satisfaction of a person in a task-oriented group is determined to a large extent by what has been termed the autonomy of the position he occupies in the information flow, greater autonomy producing higher satisfaction.

2. The above relationship operates through the mechanism of the satisfaction of a need for autonomy, a higher need producing a strong relationship.

3. To the extent that a relatively high degree of autonomy is characteristic of positions centrally located in a communication net, autonomy may be considered as mediating the observed relationship between centrality and satisfaction.

These first three conclusions roughly confirm the hypothesis of Leavitt, provided answer-getting potential in information-sharing tasks is accepted as equivalent to autonomy in activity-coordinating tasks.

The first two conclusions also parallel conclusions drawn by Morse from her correlational study of job satisfaction of white-collar workers (6). She found the characteristics of the job having the highest relationship with intrinsic job satisfaction were its variety, skill level, and reported decision-making opportunities. She also found that the strength of the last relationship varied positively with the strength of the respondents' desire to make more decisions. The similarity of these variables to the present concepts of autonomy and need for autonomy is evident.

The final conclusions are more tentative:

4. In situations of the type created for this

experiment, autonomy, contrary to hypothesis, does not also determine status judgments, while centrality apparently does.

5. This effect of centrality is particularly significant for persons low in need for autonomy.

Conclusion 4, though contrary to expectations, appears to strengthen conclusions 1 through 3, since it shows that autonomy does not constitute the full psychological meaning of centrality and thus that the autonomy-satisfaction hypothesis might well have been disconfirmed. Secondly, it casts doubt on that part of Leavitt's discussion that seems to imply that satisfaction is based on a prior perception of high status.

The role played by the need for autonomy in the centrality-status relationship, conclusion 5, suggests that under certain conditions autonomy rather than centrality might still become the major status criterion. One such condition might well be that the group have greater freedom to communicate about the problem of getting organized than obtained in this experiment. The supposition here is that while the occupant of the position of greatest autonomy is the most self-sufficient—and most powerful—member of the group, nevertheless it would take a period of perception-sharing to establish this fact and to overcome the organization-chart stereotype of peripheral members being subordinates of the central member. At any rate, the conclusion suggests that a change to the use of autonomy as a criterion for status would be facilitated by strong needs for autonomy among members.

SUMMARY

The purpose of this experiment was to test the general propositions that the autonomy of positions and the level of need for autonomy of the persons occupying the positions are significant determinants of the job satisfaction and status perceptions of members of task-oriented groups. The autonomy of a position was defined as a measure of its occupant's access to task-relevant information; it differs conceptually from centrality, which is a measure of access to communication channels, though the two may be found empirically to be positively correlated.

Forty-four college students, after answering a need questionnaire, were placed in apparent

three-person communication structures in which prewritten notes were used to create the illusions of centrality (or peripherality) and of autonomy (or dependence). The Ss were given a simple task requiring the coordination of their activities.

It was found (a) that Ss in positions of relatively high autonomy reported greater job satisfaction than Ss in relatively dependent positions, while there was no difference in reported satisfaction as between centrally and peripherally located subjects; (b) that the autonomy-satisfaction relationship was particularly strong for Ss with a relatively high need for autonomy; and (c) that, contrary to hypothesis, there was no difference in perceived status between autonomous and dependent conditions, while such a difference did exist between central and peripheral conditions.

It was concluded, first, that positional autonomy is a major determinant of job satisfaction in task-oriented groups, underlying the positive relationship between centrality and satisfaction reported in other studies. Second, the relationship between autonomy and job satisfaction is apparently mediated by satisfaction of a need for autonomy. Finally, doubt is cast on the contention that job satisfaction

results from a prior perception of a favorable status relationship.

REFERENCES

1. BAVELAS, A., Communication patterns in task-oriented groups. *J. acoust. Soc. Amer.*, 1950, **22**, 725-750.
2. CHRISTIE, L. S., LUCE, D., & MACY, J., *Communication and learning in task-oriented groups*. Cambridge, Mass., Research Laboratory of Electronics, Massachusetts Institute of Technology, 1953, (Technical Report No. 231.)
3. COHEN, A. R. The effects of individual self-esteem and situational structure on threat-oriented reactions to power. Unpublished doctoral dissertation, Univ. of Michigan, 1953.
4. GUETZKOW, H. Organizational development and restrictions in communication. Unpublished manuscript. American Documentation Institute No. 4590, 1954.
5. LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38-50.
6. MORSE, N. *Satisfaction in the white-collar job*. Ann Arbor: Univer. of Michigan Institute for Social Research, 1953.
7. SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *J. abnorm. soc. Psychol.*, 1954, **49**, 547-53.

Received May 8, 1956.

INFLUENCE OF A PARTNER ON TOLERANCE FOR A SELF-ADMINISTERED ELECTRIC SHOCK¹

DENNIS SEIDMAN, STANLEY B. BENSEN, IRWIN MILLER, AND TOR MEELAND

Human Research Unit No. 2

THE purpose of this study was to investigate the effect of one person's presence on another person's ability to endure the stress of a painful stimulus, an electric shock.

Social influences on performance under stress and the ability to tolerate stress have been widely studied. In an account of combat performance, Marshall (4) described how men became demoralized under barrage. He related this tendency to the feelings of isolation the men experience as they lie flattened on the ground. An important finding was that morale could be sustained, however, if NCOs and junior officers would move among the men, calling out their names. If this were done, units fought more effectively when the barrage was lifted.

In another report, dealing with psychiatric casualties, Mandelbaum (3) cited evidence showing that "good outcomes" are facilitated if men can be returned to their units for treatment rather than sent away from the combat zone to rest camps. This increased "good outcome" rate, in the face of the same external stresses that originally precipitated the psychiatric breakdowns, is viewed as resulting from the support provided by identification with the combat unit and its purposes.

Apart from military accounts, certain experimentally derived observations are relevant. Kelley and Thibaut (2) reviewed considerable evidence that the introduction of a partner into a situation modifies the original *S*'s response. In these experiments, *S* is given a task to complete; later a similar task is presented. At this point a partner is introduced into the situation, taking either a passive or active part in working toward the completion of the task. The changes that occur in the "together situation" as compared with the "alone situ-

ation" can be summarized as follows: (a) greater quantity of work where physical output is involved; (b) lesser quantity and quality of work where intellectual processes are involved; and (c) greater variations in work, probably due to distractions and greater tension.

While Marshall's finding that the quality of individual performance under actual combat stress can vary greatly depending on such morale factors as a man's physical and social contact with other men of his unit is unquestioned, it is not feasible in field research of this kind to focus attention on the precise social conditions which affect performance under stress. Such investigations are handicapped by limitations in research personnel and by the fact that many devices involved procedures that are too easily communicated to succeeding groups of *Ss*.

The present study, therefore, although included in a military research program concerned with stress, was not a field test, and circumstances were such that the *Ss* entered the situation without access to any knowledge of it. The test was designed to compare the tolerance for electric shock when *Ss* were tested alone and when they were tested in the presence of another *S* who apparently received the same electric shock.

METHOD

Subjects

The sample consisted of 133 men who had completed eight weeks of basic training and were due to continue with eight weeks of advanced infantry training. The age range for this group was 17 to 28, with a mean age of 20 years and 2 months. Mean intelligence was 101.35 as assessed from the Army Aptitude Area I scores, a test of general intelligence with a mean of 100 and an *SD* of 20.

Apparatus

The source of the painful stimulus was a Constant Current Electronic Stimulator (Model 228, Applegate & Co.), having a current range and a corresponding meter reading of 0 to 5 milliamperes. The stimulator delivered rectangular pulses at the rate of 60 per second with the shock control lever set for continuous current.

¹ The research reported here was conducted by the authors while they were employed by the George Washington University Human Resources Research Office, operating under contract with the Department of the Army. Opinions and conclusions are those of the authors and should not be construed as representing those of the Department of the Army.

The electrodes consisted of two silver discs coated with electrode paste and secured with an elastic band to *S*'s right calf.

Procedure

Two determinations of tolerance for electric shock were made for each *S*. In one, *S* was alone with *E*. In the other, *S* believed that a partner was also in the circuit. For both determinations, *S* himself increased the amperage to the highest level he felt he could tolerate. His instructions were to remove his hand from the current control when he reached this level. This was the signal for the *E* to interrupt the shock by means of a switch which was located out of sight of *S*.

For the condition involving a partner, *S* was told that another *S*, who had been brought in and seated next to him, would simultaneously receive a shock of exactly the same intensity. The partner, however, received no shock, although electrodes were also attached to his leg. Since *S*s were tested successively, the experimental arrangement permitted the employment of each *S* as an unwitting accomplice during the testing of the next *S*.

From 20 to 30 *S*s were tested each week for a period of six weeks. During the first three weeks, the testing order was shock alone followed by shock with partner. The testing order was reversed during the second three weeks.

RESULTS

Differences between the two shock conditions, the two presentation orders, and the six weeks of experimenting were tested for significance by an analysis of variance. This analysis also provided a means for testing whether interactions occurred between either the shock conditions and the presentation orders or the shock conditions and weeks. Since presentation order and week of testing were confounded, two alternative breakdowns of the data were made for the purpose of testing all interactions. A summary of the analysis of variance is given in Table 1.

When the differences between the shock alone and shock with partner conditions were tested by the alternative error terms, *F*s of 31.28 and 30.85 were found, indicating that the mean differences were significant beyond the .001 level. The mean current setting for shock alone was 3.13, while the mean setting for shock with partner was 3.58.

The differences between the two presentation orders were not significant, *F* being less than 1.00. Consequently, there is no reason to think that presentation order had any bearing on the results of the different shock conditions. This conclusion is substantiated by the nonsignificant shock $\times$ presentation

TABLE 1
ANALYSIS OF VARIANCE: INTENSITY OF ELECTRIC SHOCK FOR TWO EXPERIMENTAL CONDITIONS, TWO PRESENTATION ORDERS, AND SIX WEEKS OF TESTING

Source	df	Mean Square	<i>F</i>
Orders	1	21202.54	—
<i>S</i> s/Orders	131	40757.09	2.30*
Weeks	5	89018.19	—
<i>S</i> s/Weeks	127	38703.08	—
Between <i>S</i> s	132	—	—
Conditions	1	133315.88	31.28**
Conditions $\times$ Orders	1	4227.30	—
<i>S</i> s/Orders $\times$ Conditions	131	4261.56	—
Conditions	1	133315.88	30.85**
Conditions $\times$ Weeks	5	2740.30	—
<i>S</i> s/Weeks $\times$ Conditions	127	4321.18	—
Within <i>S</i> s	133	—	—
Total	265	—	—

* *P* equals .05.

** *P* equals .01 or less.

order interaction. Here also, *F* was less than 1.00.

When the overall differences among the six weeks of experimenting were tested, an *F* of 2.30 was found, showing that the weeks differed significantly at the .05 level. The mean current settings during the six weeks ranged from 2.66 in the first week to 4.12 in the second week. Since, however, the shock weeks interaction proved nonsignificant with an *F* less than 1.00, the differences among weeks plainly had no effect on the differences observed under the two shock conditions.

DISCUSSION

Under the present experimental conditions, the presence of a partner acted to raise an *S*'s tolerance level for electric shock significantly beyond that indicated when *S* was tested alone. A point of no small interest is the fact that the partner received no shock while the *S* under observation shocked himself. As could best be determined, the partner did not learn that he would not be shocked until the trial was essentially concluded. Apparently, the increase in stress tolerance was mainly a function of *S*'s perception that the stressful experience was shared by a partner. Confirmation of this point awaits further research in which the shock with partner condition can be compared with a new condition in which a partner actually does receive the electric shock.

These results may be interpreted in several ways. On the one hand, the presence of a

partner might act to raise stress tolerance by producing competitive behavior. On the other hand, the partner may provide reassurance about the ability to tolerate stress.

If the concept of stress may be legitimately generalized, an analogy may be drawn between the present study and recent investigations by Asch (1) regarding the influence of group pressures on an individual's judgment of a perceptual stimulus. Asch found that in the absence of outside corroboration, a person may experience great difficulty in resisting a prearranged group judgment that is deliberately made contrary to his own sense-testimony. The presence of one outsider, however, who substantiates *S*'s judgment greatly increases *S*'s ability to resist the pressures or stresses imposed by the group towards conformity.

SUMMARY

As part of a program of military research, the presence of a partner was explored for its effect on stress tolerance. Tolerance for electric shock was measured by having each *S* increase the shock intensity to the maximum that he could tolerate. Two observations were made for each *S*: one when he was alone, and

another when a partner apparently shared the electric shock while the *S* raised it to his own maximum tolerance level.

Analysis of data on 133 enlisted men showed that the tolerance level chosen in the condition including a partner was significantly higher ($p < .001$) than the level chosen when the men were alone with *E*. It was further shown that this result was not influenced by order of presentation or week of testing. These findings suggest that the perceived sharing of stress contributes importantly to stress tolerance.

REFERENCES

1. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
2. KELLEY, H. H., & THIBAUT, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. II. Cambridge: Addison-Wesley, 1954.
3. MANDELBAUM, D. G. Psychiatry in military society. *Hum. Organization*, 1955, **13**, 19-25.
4. MARSHALL, S. L. A. Individual summary. In S. M. Field & S. W. Davis (Eds.) *Fatigue and stress symposium, 24-26 January 1952*. Chevy Chase, Md., Operations Research Office, Johns Hopkins Univ., 10 Sept., 1952. (Tech. Mem. ORO-T-185).

Received May 7, 1956.

THE EFFECT OF THE SCHEDULE OF REINFORCEMENT ON OPERANT CONDITIONING OF A VERBAL RESPONSE IN THE AUTOKINETIC SITUATION¹

MARK SPIVAK

University of Michigan

AND JOHN PAPAJOHN

Boston University

THE purpose of this experiment is to relate the schedule of reinforcement employed in conditioning a verbal response in the autokinetic situation to the probability of the emission of the response during extinction.

Several investigators have attempted to relate different conditioning procedures to the probability of "convergence" toward a partner in experimental social settings similar to Sherif's autokinetic situation (11). Kelman (7) and Mausner (9) employed a relatively unstructured situation where it was possible to manipulate and control through differential reinforcement the reinforcement history of verbal judgments of the lengths of lines, with the dependent variable, degree of "convergence" (9) or "suggestibility" (7), being taken as the probability of occurrence of those verbal estimates or judgments which comprised *S*'s reinforced response. Certain problems raised by other experimenters (7, 9, 10) are eliminated in the present study. Kelman's claim that the establishment of a "subjective norm" (11) lessens anxiety raises the question of the final range of responses being given an indeterminate history of reinforcement, because of the reinforcement of other responses during the establishment of the final range. Also, by differentially reinforcing various *S*s, Kelman and Mausner leave their findings open to the possible alternative interpretation that convergence could be a function of the number of reinforcements received during conditioning (6).

METHOD

Apparatus. The laboratory situation was similar to that of Sherif (11). A light source was situated 5 meters from *S*. The source was an 18 watt tungsten filament bulb showing through a hole 1 mm. in diameter. Between the bulb and the hole was a sheet of ground

opal glass. There were two switches; one was used by *S* to record when he saw the light begin to move; the other was operated by *E* to turn the light on and off. The power source and *E*'s switch were part of the equipment used in an introductory course in psychology (4). The *S* sat at a small table on which the switch was placed. When *S* was in the experimental room no other light besides that from the light source was present.

Subjects. The *S*s were 18 undergraduate volunteers from a laboratory section in general psychology. All the *S*s were in their second term of psychology and had had some experience in conditioning and extinguishing procedures in student experiments with animal subjects.

Method. The *S*s were divided randomly into two groups of nine. Each *S* was told that he was to participate in an experiment designed to measure the variability of college students in judging lengths of lines. The following instructions were given:

You will be taken into the experimental room with your eyes closed. The *E* will place your hand on a table. Work your way to the end of the table where you will find a chair. Sit down on the chair. Find a switch on the table. With one hand hold the switch steady and use the other hand to operate the switch. Keep your eyes shut until *E* tells you to open them. After you open your eyes, *E* will say "ready." A few seconds later a light will go on. When you see the light begin to move, close the switch. Shortly after this the light will go off. At this point call out the distance the light moved in inches. The light will move any distance between 0 and 60 inches. When you guess the correct distance *E* will call "RIGHT"; when you guess the wrong distance *E* will not say anything, and will go on to the next trial. Before the experiment begins in earnest, you will have five trials to become acquainted with the apparatus. Any questions?

The *S*s were led one at a time into the experimental room with their eyes shut, in order to insure the unstructured qualities of the situation. When *S* was seated *E* left the room and shut a light-proof door which separated *S* from the room in which *E* operated his switch and recorded each estimate as it was given.

During the five practice trials, for half the *S*s in each group, *E* recorded the estimates and selected a mean as the point to be reinforced in subsequent presentations of the light. For the other half of the *S*s the first estimate after the practice preliminary trials was chosen as the response to be reinforced. Hereafter,

¹ This study was carried out while the authors were graduate students at Columbia University.

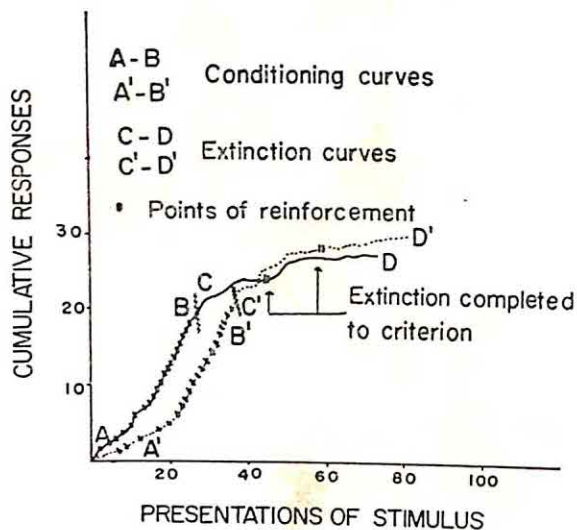


FIG. 1. TWO TYPICAL CONDITIONING AND SUBSEQUENT EXTINCTION CURVES OF Ss UNDER A REGULAR REINFORCEMENT SCHEDULE

"response"—as distinguished from "estimate"—denotes only that estimate chosen to be reinforced, whether it occurred in the conditioning or extinction phase.

One group of nine Ss was reinforced on a Regular Reinforcement schedule (RR). The other group of nine Ss was reinforced on a Variable Interval reinforcement schedule (V-I), i.e., under conditions of partial reinforcement. Five Ss in each group were reinforced for every estimate of the same length as the first estimate after the five practice trials. The four remaining Ss in each group were reinforced when they gave an estimate close to the mean of the five preliminary practice estimates.

For the RR group the first 20 responses were reinforced and then extinction was introduced until a criterion was reached of 10 estimates without the designated response. For each S in the V-I group, the first 5 responses were reinforced regularly, after which 15 reinforcements were given on a V-I schedule. After 20 reinforcements had been given, extinction was introduced. Since no S was extinguished in this group, extinction training was continued for a minimum of 80 judgments.

When the presentations of the light were completed, S was told to close his eyes and was led back to E's office, where he was questioned on his impressions of the experiment.

Several authors (3, 14) have discussed the rationale of treating verbal estimates as magnitudes. The general opinion is that although these estimates are "number responses" the relation of an ordering procedure of responses of this nature to various stimuli reveals the relation not to be linear. It may be assumed, however, that by intensive training procedures, it would be possible to obtain number responses from Ss which would have upon ordering a 1:1 relation with the ordering of the actual linear stimuli upon the presentation of which the responses were emitted. On this assumption, operations of arithmetic are acceptable since a zero point and equal intervals are assumed.

The first author, who was the Ss' laboratory assistant, alternated with the second author, who was unknown to the Ss, in conducting the experiment.

RESULTS

Effects of regular reinforcement on conditioning and extinction. After being reinforced a few times Ss conditioned under RR began responding maximally, i.e., they emitted the designated response at every presentation of the light for the duration of the conditioning period. When extinction was introduced, there was a sharp drop in responding, until the criterion of extinction was reached, i.e., until 10 estimates without a response were emitted. Two representative individual curves are shown in Fig. 1. Each of the nine Ss in this group manifested the same behavior with regard to (a) achieving the maximum rate of responding in a short time; (b) extinguishing to the criterion.

Effects of variable interval schedule of reinforcement on conditioning and subsequent extinction. For the group of Ss conditioned under a V-I schedule, Fig. 2 shows that in all Ss a stable strength of response was built up in the conditioning period. When extinction was introduced after the 20 reinforcements had been

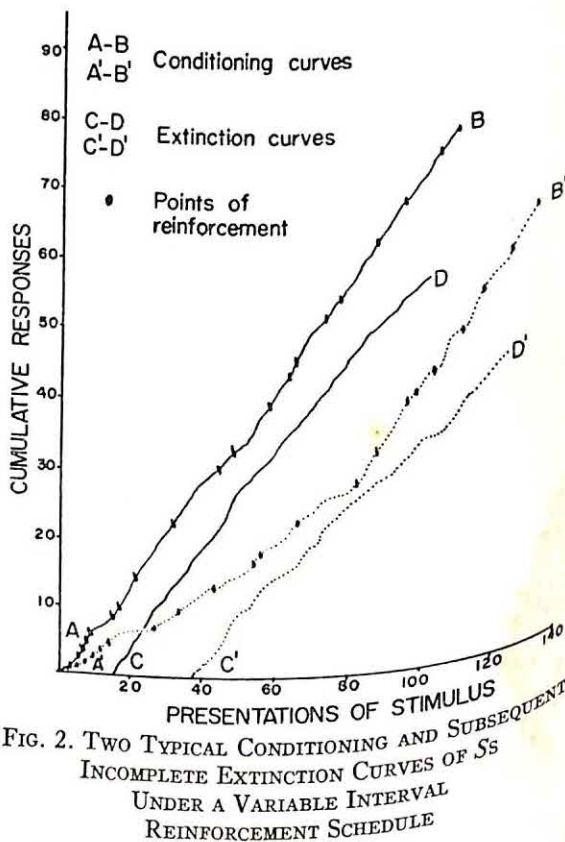


FIG. 2. TWO TYPICAL CONDITIONING AND SUBSEQUENT INCOMPLETE EXTINCTION CURVES OF Ss UNDER A VARIABLE INTERVAL REINFORCEMENT SCHEDULE

given the emission of responses remained high. No *S* met the criterion of extinction within the limits of time set for the experiment. However, each *S* in the V-I group was recorded as emitting more than a total of 80 judgments without the extinction criterion being reached.

In interview only two *Ss* in the RR group and one *S* in the V-I group reported that they thought they had been subjected to conditioning and extinguishing procedures. There was nothing in their conditioning and extinction curves that distinguished them from the other *Ss* who did not show insight into the experimental procedure.

Several *Ss* reported that when they were not sure how far the light had moved, they tended to give as an estimate the reinforced response. During extinction for *Ss* reinforced on the V-I schedule it was also reported that at times the response was given almost automatically as soon as the light went on.

The average variability of the estimates around the reinforced response, disregarding direction, was computed by the Vincent method for each group during extinction. Figure 3, which shows the deviation from the reinforced response as a function of time during extinction, indicates that the variability of the RR group increased more as extinction continued and at a steeper rate than in the V-I group.

Similar results were obtained without assuming the validity of using "number re-

sponses" in a quantitative manner by counting the number of verbal emissions other than the designated responses in each fifth of the extinction period. Figure 4 indicates the RR group gave a greater number of different verbal emissions during extinction than the V-I group, indicating greater learning in a nonrewarding situation. A greater percentage of the total verbal emission were other than responses in each fifth of the curve for the RR group than for the V-I group.

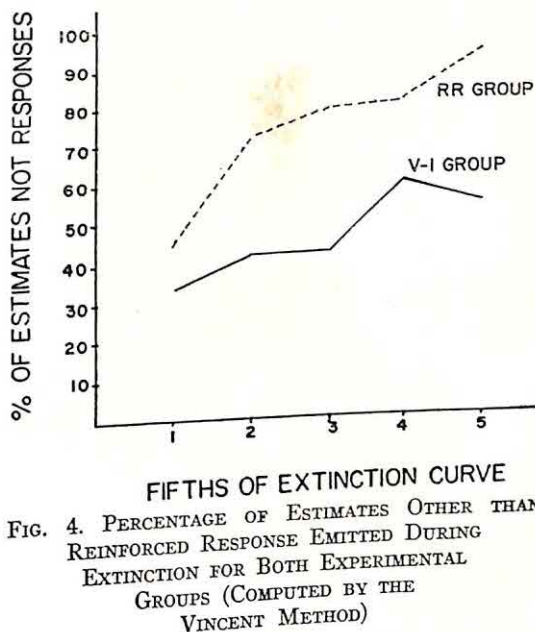
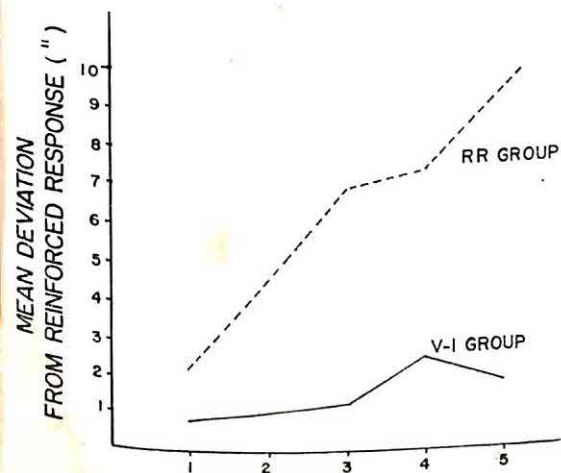


FIG. 4. PERCENTAGE OF ESTIMATES OTHER THAN REINFORCED RESPONSE EMITTED DURING EXTINCTION FOR BOTH EXPERIMENTAL GROUPS (COMPUTED BY THE VINCENT METHOD)



FIFTHS OF EXTINCTION CURVE

FIG. 3. THE MEAN DEVIATIONS AROUND THE REINFORCED RESPONSE, EMPLOYING THE VINCENT METHOD, DURING EXTINCTION, FOR BOTH EXPERIMENTAL GROUPS

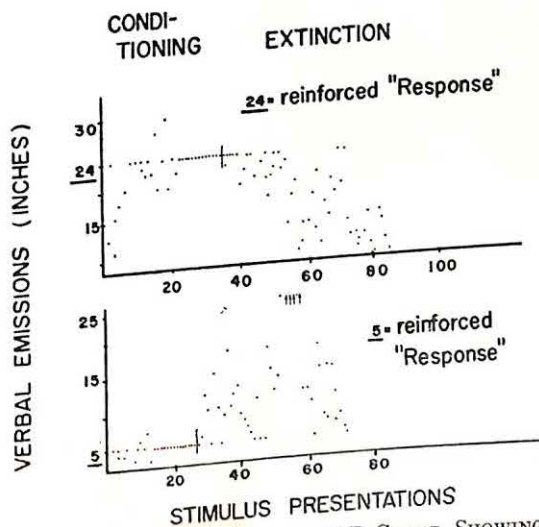


FIG. 5. DATA OF TWO *Ss* IN THE RR GROUP, SHOWING VERBAL ESTIMATES OF DISTANCE LIGHT MOVED AS A FUNCTION OF SUCCESSIVE PRESENTATIONS OF THE STIMULUS (LIGHT)

CONDITIONING

EXTINCTION

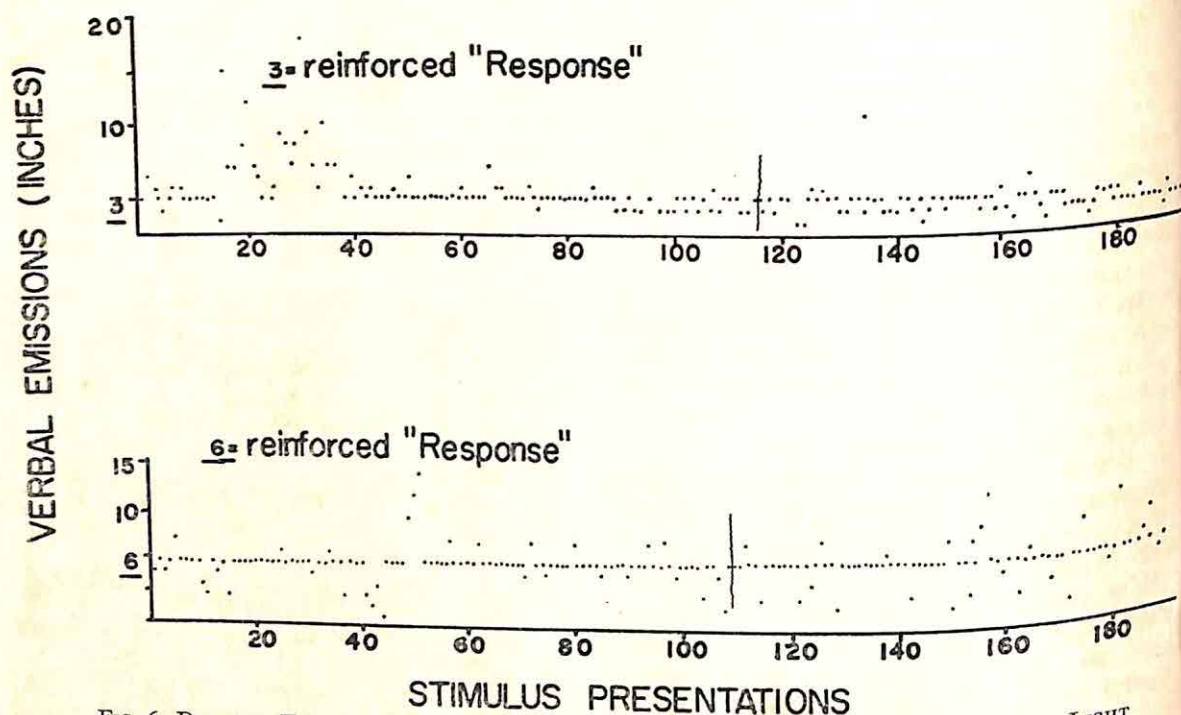


FIG. 6. DATA OF TWO Ss IN THE V-I GROUP, SHOWING VERBAL ESTIMATES OF DISTANCE LIGHT MOVED AS A FUNCTION OF SUCCESSIVE PRESENTATIONS OF THE STIMULUS (LIGHT)

Graphic representations of the numerical data for the two illustrative Ss from the RR and the V-I groups are provided in Figs. 5 and 6, respectively. Each point represents the S's judgment for a particular presentation of the stimulus. It will be noticed in Fig. 5 that during extinction, the verbal emissions vary widely from the formerly reinforced response and few verbal emissions are responses, while in Fig. 6 the verbal emissions during extinction remain close to the formerly reinforced response, and many verbal emissions are responses.

DISCUSSION

In consonance with previous investigations (6, 12, 15) the type of reinforcement schedule used in conditioning is directly related to resistance to extinction, providing confirmatory evidence that the principles of reinforcement found in the patterning of response emissions of infrahuman organisms, also hold with respect to verbal behavior.

A possible explanation is suggested by Skinner (13) for the fact that it proved impossible to extinguish Ss conditioned under a V-I schedule within the time limits available

for this experiment. He implies in his statement concerning "generalized reinforcement" of *tacts* (verbal responses) that the omission of direct negative reinforcement may allow generalized secondary reinforcements to carry over to the following S-R situation.

The present results indicate that variability of verbal behavior is greater during extinction following RR conditioning than during extinction following V-I conditioning. Consistent with the present findings is a recent report (5) that an RR group of rats affected a reversal in discrimination response significantly more rapidly than a V-I group. In a study in progress the present authors are finding that reconditioning of Ss previously trained in the procedures in this experiment on the two schedules but not extinguished bear out the observations of Grosslight *et al.* (5), i.e., on reconditioning on a V-I schedule after conditioning to a V-I schedule to a different response the S requires a significantly greater number of reinforcements to condition and his extinction period is shorter than Ss who were reconditioned on a V-I schedule after first being conditioned on a RR schedule to a different response. In other words, variability in response seems to be as-

sociated with faster learning in situations where the contingency of reinforcement is changed.

In relation to the experimental tradition of Sherif (10), Kelman (7), or Mausner (9) among others, this study appears to demonstrate the possibility of giving *S* a completely determined history of reinforcement with respect to a certain response, which then controls his behavior in an orderly and predictable manner. The variable of reinforcement history could well be introduced into social psychological experiments using the autokinetic effect.

With the measure of the strength of the reinforced verbal response used in this experiment, i.e., resistance to extinction, many of the difficulties encountered with convergence or suggestibility are avoided. Individuals vary greatly in the degree to which they resist social influence on their judgments (1). Kelman (7) tried to control the "personality" factor by ascertaining "ascendancy-submission" through the Guilford-Martin Inventory (8). These traits were evaluated as independent variables together with the experimentally introduced success and failure. Bray's attempt (2) to relate the amount of convergence by a judge towards a Jewish, Negro, or a white Protestant partner to the amount of prejudice in the judge indicates that the interaction factor must be considered in any social situation. The consistent results of the randomly chosen *Ss* under each schedule of reinforcement in this experiment, however, demonstrate that the problems raised by personality and social interaction variables are largely eliminated by the present procedure.

SUMMARY

An experiment employing the autokinetic effect investigated the effect of two different schedules of reinforcement on the conditioning of verbal behavior. Nine undergraduate psychology students were given 20 reinforcements of "Right" under a Regular Reinforcement schedule for emitting verbally the estimate designated by the experimenter as the response. Nine others were given similar reinforcement under a Variable Interval schedule. Extinction was then carried out until a criterion of 10 estimates other than the designated response was reached.

The RR group reached the extinction criterion within the limits of 10 to 35 verbal emissions, although no *S* in the V-I group was ex-

tinguished within the limits of 10 to 80 verbal emissions. The results are consonant with reinforcement principles.

It is argued that the design of this experiment corresponds to social psychological experiments on perception and motivation, but eliminates several difficulties in previous studies. It is suggested that the variable of a determinate reinforcement history can fruitfully be introduced in such experiments.

REFERENCES

1. ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (Rev. Ed.) New York: Holt, 1952. Pp. 2-10.
2. BRAY, D. W. The prediction of behavior from two attitude scales. *J. abnorm. soc. Psychol.* 1950, **45**, 64-84.
3. CRONBACH, L. J. The application of statistical methods to Rorschach scores. *Psychol. Bull.* 1949, **46**, 393-430.
4. FRICK, F. C., SCHOENFELD, W. N., & KELLER, F. S. Apparatus designed for introductory psychology at Columbia College. *Amer. J. Psychol.* 1948, **61**, 409-414.
5. GROSSLIGHT, J. H., HALL, J. F., & SCOTT, E. Reinforcement schedules in habit reversal—a confirmation. *J. exp. Psychol.* 1954, **48**, 173-174.
6. KELLER, F. S., & SCHOENFELD, W. N. *Principles of psychology*. New York: Appleton-Century-Crofts, 1950.
7. KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.* 1950, **45**, 267-285.
8. MARTIN, H. G. The construction of the Guilford-Martin Inventory of Factors GAMIN. *J. appl. Psychol.* 1954, **29**, 298-300.
9. MAUSNER, B. The effect of previous reinforcement on interaction of observer pairs. Unpublished doctor's thesis, Columbia Univer., 1950.
10. SHERIF, M. Group influences upon the formation of norms and attitudes. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (Rev. Ed.). New York: Holt, 1952. Pp. 249-262.
11. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.* 1935, No. 187.
12. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century-Crofts, 1938.
13. SKINNER, B. F. William James lectures on verbal behavior. Harvard University, 1948. In F. S. Keller & W. N. Schoenfeld, *Principles of psychology*. New York: Appleton-Century-Crofts, 1950.
14. STEVENS, S. S. On the problem of scales for the measurement of psychological magnitudes. *J. unified Sci.* 1939, **9**, 94-99.
15. WIKER, E. L. Extinction of a partially and continuously reinforced response with and without a rewarded alternative. *J. exp. Psychol.* 1953, **46**, 255-260.

Received May 10, 1956.

AN EXPERIMENTAL STUDY OF CROSS-CULTURAL NORM CONFLICT¹

JOHN R. P. FRENCH, JR. AND ROBERT B. ZAJONC

Research Center for Group Dynamics, University of Michigan

NORM conflict exists when an individual is subject to contradictory behavioral prescriptions. Since norms have the character of pressures or forces toward conformity, an investigation of norm conflict must begin with the examination of the sources and nature of these pressures. With regard to the source of pressures two types of norm conflict can be distinguished: *intragroup* and *intergroup* norm conflict. The first type occurs when a given group has norms prescribing different or contradictory kinds of behavior for a particular situation. All members of the group who find themselves in that situation experience conflict. In the instance of intragroup norm conflict examined by Stouffer (9), students were in conflict between norms governing relations and loyalty among friends and those governing the responsibilities of an honest proctor.

Intergroup norm conflict exists when the individual is a member of two or more groups having different or contradictory norms for the same situation. Instances of this category are: the conflict of American adolescents born of immigrant parents as studied by Child (1); the adjustment of Navaho veterans examined by Vogt (10); the problems arising from cultural diffusion such as those explored by Spindler and Goldschmidt (8); and Rommetveit's studies of the conflicting religious norms prescribed for a child by his parents vs. his peer group (6). Another example of intergroup norm conflict is readily observable today in the area of cross-cultural education (7, 11). Usually the foreign student maintains some degree of membership in two cultures. By virtue of his membership in the home country one set of norms demands conformity from him, and by virtue of his temporary membership in the host country another set exercises

such pressures. Often the two norms stand in clear conflict with one another.

The crucial theoretical factor distinguishing between intra- and intergroup norm conflict is the role of group membership in conformity. Suppose, for example, we place the individual in a norm-conflict situation and are interested in predicting to which norm he will conform. In the case of intragroup norm conflict, group membership does not influence his choice since forces deriving from it act with equal magnitudes on conformity to both norms. In this instance the important independent variables are likely to be found in individual dynamics, situational factors, or the normative pressures themselves. In the case of intergroup norm conflict, however, the examination must begin with group membership. Given different strengths of membership we should expect corresponding differences in the pressures toward conformity. If the individual's membership in Group A is stronger than that in Group B there should be stronger forces on him to conform to the norm of Group A than to the norm of Group B. An attempt is made in this study to predict the individual's behavioral choice under conditions of intergroup norm conflict by analyzing the role of group membership in its resolution.

THEORETICAL FORMULATION OF THE PROBLEM

Our theoretical formulation is based on Lewin's conceptualization of the quasi-stationary equilibrium (2, 5), and on French's theory of social power (3). For the present purposes *norm* is defined as a set of forces exerted by a given group toward conformity along a specific behavioral dimension to a given standard or ideal (i.e., a specific point or range of points on that dimension) for group members who are in situations requiring behavior along that dimension. We shall assume that the forces toward conformity at the ideal are equal to zero, and that they increase monotonically with the distance from the ideal. This increase may reach an asymptote at a certain point.

¹This work was sponsored by the Social Science Research Council as part of a larger program on cross-cultural education. The authors are indebted to M. Brewster Smith, Rose Goldsen, Richard T. Morris, Claire Seltiz, and other members of the program for their valuable suggestions during the planning of the study.

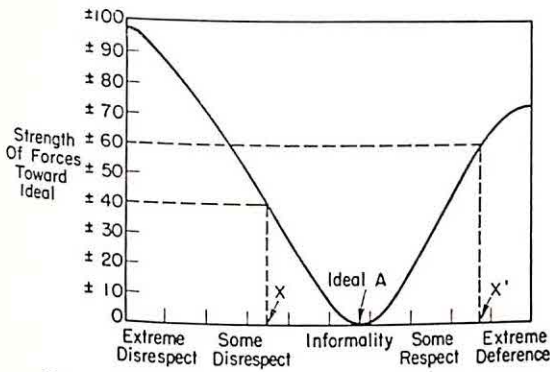


FIG. 1. FORCEFIELD REPRESENTING A NORM

The direction of the forces is assumed always to be toward the ideal.

Norms, having the formal properties of quasi-stationary equilibria, may be treated as forcefields represented in a two-dimensional space (5, p. 199-237). Let the abscissa be the given behavioral dimension, such as "showing respect to professors." Let the ordinate represent the strength of forces to conform to the behavioral standard or ideal, as seen in Fig. 1. In the figure, the magnitude of the forces on the left of the ideal (i.e., exerting pressures toward greater deference) is denoted by positive numbers, and the magnitude of the forces on the right of the ideal (i.e., exerting pressures toward less deference) by negative numbers. As the behavior departs from the ideal the forces acting on the group member to conform increase. Thus, a member who behaves at point x would experience a force equal to +40 in the direction of showing more respect, while a member behaving at point x' a force equal to -60 in the direction of less deference.

Suppose the individual is subjected to forces deriving from two conflicting norms and we are interested in discovering how he resolves his conflict. The two norms can be represented on the same dimension as shown in Fig. 2. The individual who behaves at the point x is now under the impact of two forces: one equal to +40 in the direction of Ideal A, and another equal to -60 in the direction of Ideal B. For the purpose of further derivation we shall assume that the direction and magnitude of the resultant force acting on the individual can be obtained by adding the two magnitudes. In the above case the resultant is equal to -20, and the direction, the number being negative, is toward Ideal B.

If we assume that behavior takes place in the direction of the resultant force, we can derive our first hypothesis: *behavior in a situation involving pressures to conformity to two different normative ideals occurs at the intersection of the two corresponding forcefields.* Of course, other factors, such as the extent of interiorization of the norms, personality, and situational variables, also influence conformity in a conflict situation. In testing this general hypothesis we assume that these factors are held constant.

When the two norms derive from the same group, the shape of the curves does not depend on group membership. If, however, they stem from two different groups, the strength of the forces and the shape of the curves vary as a function of strength of membership in the two groups. The stronger the ties of membership the greater is the power of the group in obtaining conformity from its members. Thus, in the case of intergroup norm conflict group membership becomes an important factor in determining the position of the intersection of the two forcefields and the resultant behavior.

The effects of group membership on conformity, however, are not constant across all situations, as Lewin recognized in his concept of the potency of social groups (4, p. 85). In some cases the individual may behave more as a member of one group than as a member of another, while in others both group memberships may exert equal pressures. We therefore must take into account the *situational potency of membership*, defined as *the degree to which a given group membership is aroused in a particular situation relative to the degree to which his membership in other groups is aroused.* It should be noted that situational potency of a

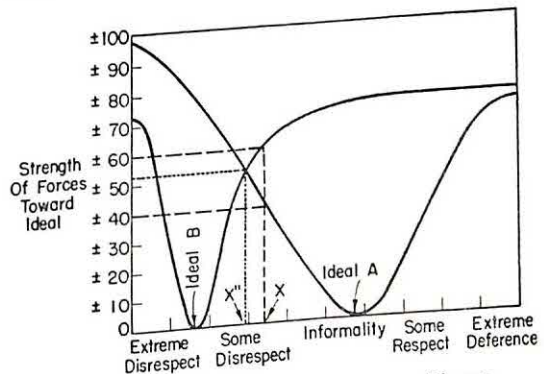


FIG. 2. FORCEFIELDS REPRESENTING NORMS IN CONFLICT

given membership is always relative to situational potencies of other memberships. All potencies in a given situation, therefore, add to unity and each particular potency is represented by a fraction of unity.

A second hypothesis regarding the resolution of intergroup norm conflict can be derived by assuming that the component forces of the forcefield representing a norm of a given group increase monotonically with the situational potency of membership in that group. It follows that when potency of membership is manipulated in a situation of intergroup norm conflict, *the point of the intersection between the two norms and thus the resulting behavior is shifted in the direction of the normative ideal of the group whose potency is increased, and away from the ideal of the group whose potency is decreased.*² In other words, if the individual is in a situation governed by conflicting norms of two groups, he tends to deviate from the norm of the group that has less situational potency for him, and to approach the normative ideal of the group that has high situational potency.

METHOD

This study employed a combination of laboratory and field methods. Cross-cultural problems have rarely been examined in a laboratory setting, probably because of the difficulties arising in attempts to reproduce cultural forces experimentally. The laboratory method promises precision and rigor, but it often does so at the cost of departing from the practical significance of the problem. The procedures employed in field studies make contact with phenomena of practical importance, but they seldom allow the precise controls which are crucial in measuring variables in their pure uncontaminated form. In the present investigation, we attempted to bring "real" cultural forces into the laboratory and to examine their effects under controlled conditions. The subjects were all foreign students, with their own sets of norms, values, and patterns of behavior. In the laboratory measurements of norms were obtained and the Ss were faced with a "test" situation designed to manifest the operation of cultural forces under conditions of norm conflict.

The study required individuals who hold simultaneous memberships in two distinct groups with different norms. To meet this requirement, Indian students studying in the United States were used as Ss.³ We assumed that these individuals maintain temporarily two group memberships, Indian and

American, and are therefore subject to two different sets of norms. On the basis of a preliminary investigation which disclosed that considerable differences exist between Indian and American norms of deference to professors, this area of potential conflict was selected for experimental purposes.

During the course of the experiment measurements were obtained regarding the Indian and American deference norms as perceived by the Ss. From these measurements predictions were made about the resolution of norm conflict. The predictions were tested by confronting the Ss with a "professor," while the situational potency of membership was systematically varied. High potency of Indian membership was produced by certain instructions and by having Ss interact with an Indian "professor," whereas high potency of American membership involved interaction with an American "professor." Both "professors" were actually paid participants trained for the purposes of the experiment.⁴ During the experimental sessions the Ss discussed controversial items with each "professor," separately. The interaction took place over an intercommunication system, with the parties not being able to see each other. The entire session was tape recorded, and measures obtained from the recordings were correlated with the predictions.

The experimental sessions. All the experimental sessions were held individually. Upon arrival S was greeted by the experimenter and told that the purpose of the study was to discover the different ways in which students work with their professors in various countries. He was further told that he would be working with two professors who were also experimental subjects as he was. One of these was said to be an Indian professor, currently visiting the United States, and the other an American professor at the University, both being professors of psychology.

The S was told that for the purposes of scientific rigor all the communication between him and the "professors" would be by means of an intercom system. He was given a chance to try the workings of the system, which gave him an opportunity to get acquainted with the "professors," who remained in another room during the entire session.

The conversation between the S and the "professors" was held in English. As mentioned above, both were paid participants. The American "professor" was portrayed by an American graduate student, the Indian "professor" by a Pakistani graduate student who lived in India most of his life. The latter spoke with a clearly detectable Indian accent. The Ss never doubted that they were interacting with bona fide American and Indian professors. During the "getting acquainted" period many Ss invited the Indian "professor" to spend some time together after the experiment was over.

The task. The task consisted of discussing four controversial items designed to increase the salience of deference norms. The discussion always involved the student and one of the "professors." Below is a sample item of discussion.⁵

⁴ They were M. Sulaiman Kakli and Donald M. Wolfe, who also served as observers and statistical analysts.

⁵ The entire set of discussion items and the instru-

² This statement is similar to Rommetveit's hypothesis about compromise between conflicting norms but without the qualification that too great a discrepancy will lead to a forced choice and complete conformity to one norm and rejection of the other (6, pp. 68-69).

³ They were 44 male Indian students enrolled in the University of Michigan during the 1954-1955 academic year.

It is a matter of fact that some people have better memory than other people. But it is a matter of opinion whether these differences are due to hereditary or environmental influences. That is, some psychologists believe that good or bad memory is given at birth, and other psychologists believe that people acquire good or bad memory later in life. What is your opinion?

1. Heredity is the only factor.
2. Heredity is most important in determining memory; environment plays a minor role.
3. Heredity is considerably more important than environment.
4. Heredity is somewhat more important than environment.
5. Heredity and environment play about equal parts in determining memory.
6. Environment is somewhat more important than heredity.
7. Environment is considerably more important than heredity.
8. Environment is most important in determining memory; heredity plays a minor role.
9. Environment is the only factor.

The items were constructed with the intent of making each position equally defensible. The *S* and the "professor" were given one item at a time to discuss. The "professor" allowed the *S* to select a position on the scale, and then selected a position three points away from that of the *S*. Each of the "professors" had two lists of four arguments for each item, one to be used if the *S* happened to choose the "heredity" side of the question, the other when he chose the "environment" side. The order of the items was varied at random among the "professors," and the sequence of the discussants was also rotated, so that for one *S* the American "professor" would be first, and for the following *S* the Indian "professor" would be first.

The paid participants were instructed to stay as much as possible within the wording of the prepared arguments. During the training sessions an attempt was made to equate the behavior of the "professors" on such dimensions as forcefulness of voice, speed of verbalization, timing of arguments, and clarity. On the basis of rating their behavior by inspection of pretest cases fair confidence is justified that reasonable uniformity of behavior on these dimensions was obtained.

Observation measures. For the purposes of the experiment deference behavior was defined as behavior in a situation involving sub- and superordinate social positions, in which the occupant of the subordinate position overtly acknowledges the superordinate position of the other, or his own subordinate position. It was assumed that the readiness with which the student accepted the "professor's" arguments, the amount of formal address, politeness, frequency of acknowledgment of the "professor's" status, knowledge, or power,

were positive indicators of deference, while rejection of the "professor's" arguments, attempts to convince him, impoliteness, informality, and emphasis on *S*'s own knowledge, were negative indicators of deference. Observations were made from the tapes by two observers independently. Deference behavior was measured on the basis of the above indicators, which were scored on six three-point scales. Each *S* was given a final score of "observed deference," based on the index computed from the scores he obtained on each of the scales.

The over-all interobserver reliability was .80, obtained from correlating the total deference scores for each *S* resulting from independent observations from the tape recordings by two observers. For an additional check on reliability, three *S*s selected at random were observed by both observers independently. For each *S* comparisons were made to determine the interobserver agreement for each unit of deference behavior in terms of the six three-point scales measuring deference. Agreement for the three *S*s was respectively 92%, 87%, and 89%, in each case significantly higher than the chance expectation of less than 6% for six three-point scales.

The manipulation of situational potency of membership. The *S* was told at the beginning of the experiment that the researchers were interested not so much in the student himself as in the student and professor as a "working team," and that the object of the study was to compare such "working teams" across various nationalities.

The *S* was further told that during one half of the session you will be a member of the "Indian team" and during the other a member of the "Indian-American team." When working with the Indian professor, please try to behave as you would behave in India, or as you did when you went to school back home. When working with the American professor behave as you normally do in the United States. Remember, we are more interested in the teams from the different countries than in individuals.

The success of this induction was tested by means of the following questions asked postexperimentally: "While talking with the Indian professor did you feel more like an American student, or more like an Indian student?" and "While talking with the American professor did you feel more like an American student, or more like an Indian student?" The results are shown in Table 1. It can be seen that the induction was successful in only one-third of the cases under the Indian condition, and in only about one-fourth of the cases of the American condition. Thus significantly more *S*s failed to respond to the induction positively. The tests of our hypotheses would have been more accurate if the manipulation of situational potency of membership were more successful.

Postexperimental measurements. After the experimental session the *S*s were administered three instruments. The first was designed to obtain forcefields on the American norm of deference as perceived by the Indian students, and the second to obtain forcefields on the Indian norm of deference. The third instrument was designed to test the assumptions made in this study, and to obtain measures of the success with which potency had been manipulated.

TABLE 1
SUCCESS OF POTENCY MANIPULATION

Response	Condition	
	Indian	American
Felt more like Indian Student	15*	6
Felt more like American Student	4	10*
Both or neither	25	28
	$\chi^2 = 4.46$	$\chi^2 = 13.10$
	$p < .05$	$p < .001$
	$df = 1$	$df = 1$

* The tests in each treatment compare the frequencies which responded to the induction positively (identified by asterisks) and negatively. Thus, for instance, in the Indian condition the frequency of Ss who "felt more like Indian students" was compared with the frequency of Ss in the two remaining categories combined.

The forcefields were measured in terms of the following operational definitions: *The strength of the forces acting on a member who occupies a given point on the behavioral dimension governed by the norm to conform to the normative ideal corresponds to the inverse of the proportion of group members he perceives as occupying the same point.* The scales developed on this basis consisted of ten examples varying from extremely high to extremely low deference.⁶ In the first instrument the S was asked to state what percentage of American students behave in a manner similar to that shown in the example, and to rate each example on a 31-point scale (-15 to +15) in terms of the amount of deference it connotes in America. Thus we measured for each S, the sizes of the intervals between examples rather than assuming an equal interval scale which would be equally applicable to all Ss. In the second instrument the same examples were used and Ss performed similar tasks, except now with reference to Indian students and Indian professors.

RESULTS

Testing Major Hypotheses

Our first general hypothesis states that *behavior in a situation involving pressures to conformity to two different normative ideals occurs at the intersection of the two corresponding forcefields.* This hypothesis predicts normative behavior for the case where the two norms are held by the same group (intragroup norm conflict) as well as for the case where they are held by different groups (intergroup norm conflict), but only under conditions of equal situational potency of membership. In the present case of intergroup norm conflict the hypothesis implies that Ss who manifest equal situational potency of membership in the Indian and American groups should be

⁶ See Footnote 5.

TABLE 2
CORRELATION BETWEEN "PREDICTED" AND "OBSERVED DEFERENCE" UNDER CONDITIONS OF EQUAL AND UNEQUAL POTENCY

Condition	Potency	
	Equal	Unequal
Indian	.31	.06
American	.22	-.32

more likely to behave around the point of intersection between the two respective forcefields than those who manifest unequal potency. Both groups of Ss were identified on the basis of questionnaire responses to the manipulation of potency.

To test the hypothesis, a prediction was made for each S about his expected deference toward the American and toward the Indian "professors." The predicted deference scores were obtained in the following way. For each S two curves, similar to those shown in Fig. 2, were constructed on the basis of his perception regarding the distribution of deference behavior in India and America respectively. The abscissa represented the amount of deference shown, varying from extreme disrespect to extreme deference. In accordance with the operational definition of normative forces (p. 11), the ordinate represented the perceived proportion of the relevant population occupying the various points on the abscissa. Once the two curves were available the point on the abscissa where the curves intersected was readily obtained. These points represented our predictions regarding the amount of deference expected of the Ss in the test situation. Index scores of observed deference constructed from the observational categories were obtained by observers from the tape recordings of the experimental sessions. To test the hypothesis "predicted deference" was correlated with "observed deference" for Ss with equal and unequal potencies of membership.

The results are shown in Table 2. The correlation coefficients were somewhat higher for Ss with equal than with unequal situational potencies of membership, although the differences are significant only under the American condition ($p < .05$). In spite of the evidence obtained for a systematic relationship between the two variables, however, the existence of a correlation in itself does not

demonstrate that behavior occurred precisely at the point of intersection.

The second hypothesis holds that when potency of membership is manipulated in a situation of intergroup norm conflict, *the point of intersection between the two norms and thus the resulting behavior is shifted in the direction of the normative ideal of the group whose potency is increased and away from the ideal of the group whose potency is decreased.* The Indian norm calls for higher deference to professors than the American norm. The mean perceived Indian normative ideal was 11.09, while the mean perceived American ideal was 8.30, this difference being significant at the .05 level.⁷ By hypothesis, observed deference should therefore be higher in the condition where Indian potency was heightened than in the condition where American potency was heightened. The results tend to support this hypothesis. In the Indian condition mean observed deference was 153.6, in the American condition 102.6. The difference between these two means is significant at the .05 level.

It will be recalled, however, that the manipulation of potency was not entirely successful. Means of observed deference were therefore computed for Ss who "felt more like American students," for those who "felt more like Indian students," and for those who "felt like neither or both." By hypothesis, Ss with high American (and therefore low Indian) potency should score low on observed deference, Ss with high Indian (and therefore low American) potency should score high, and Ss with equal potencies should obtain medium scores on observed deference.

Table 3 presents the results of this analysis. Although the results are all in the predicted direction analysis of variance failed to show significant F ratios ($p < .09$ for the Indian condition, $p < .15$ for the American condition), as a result of small N s and unexpectedly high variance in observed deference. For instance, SD for the total scores in the Indian condition was 246.6 and in the American condition 193.4. If comparison is restricted to Ss who responded positively to the manipulation of

⁷ The scales of deference ranged from -15 to +15 representing a dimension going from extreme disrespect to extreme deference. To obtain a measure of the normative ideal Ss were asked to indicate how much deference "should" be shown to professors in each country.

TABLE 3
OBSERVED DEFERENCE FOR SUBJECTS WITH VARYING
SITUATIONAL POTENCY OF MEMBERSHIP

Condition	Mean Observed Deference			
	High Indian-Low American Potencies	Equal Potencies	Low Indian-High American Potencies	All Subjects
Indian	227.3 (15)*	116.5 (25)	109.1 (4)	153.6 (44)
American	128.4 (6)	100.8 (28)	82.1 (10)	102.6 (44)

* Figures in parentheses indicate the N s.

TABLE 4
PREDICTED DEFERENCE AND FELT
CONFORMITY PRESSURES

Felt Conformity Pressure	Mean Predicted Deference	
	Indian	American
High	8.79	7.26
Low	6.53	5.02
$t = 1.46$	$p < .07$	$t = 1.35$
		$p < .08$

TABLE 5
OBSERVED DEFERENCE AND FELT
CONFORMITY PRESSURES

Felt Conformity Pressure	Mean Observed Deference	
	Indian	American
High	240.77	157.93
Low	44.28	43.20
$t = 2.34$	$p < .01$	$t = 2.90$
		$p < .01$

potency, a significant difference is obtained. The mean observed deference for Ss with high Indian potency in the Indian condition was 227.3, and for Ss with high American potency in the American condition 82.1. This difference is significant at the .05 level ($t = 2.12$).

Testing Assumptions

Several assumptions were made in deriving hypotheses, as well as in constructing measuring instruments. Results testing these assumptions are presented below.

Difference in deference norms. To be suitable for the present study, a dimension of behavior was required in which considerable differences in prescribed behavior exist between the Indian and the American cultures. The comparison between the modes of the forcefields of both cultures has shown significant results. These modes represent points on the abscissa perceived as being occupied by the highest proportion of the relevant population. The

average of these modes was 8.07 for India, and 5.83 for America, this difference being significant at the .001 level. Sixty-four per cent of the sample gave the Indian mode as higher than the American mode, while 18 per cent showed the opposite trend and 18 per cent no difference between the modes. The mean modes of the forcefields are consistently lower than the reported normative ideals⁸ (i.e., 11.09 and 8.30 for India and America respectively), indicating that the Ss perceived the majority of Indian and American students as showing less deference than required by the respective ideals. The differences between the normative ideals and the mean modes were significant at the .001 level for the Indian norm, and at the .01 level for the American norm.

Experience of conformity pressures and deference. It was assumed that the forcefield curves represent normative pressures as they exist for each individual. In order to test this assumption the "predicted deference" scores obtained from these curves were related to Ss' report regarding felt conformity pressures, obtained in the postexperimental interview. The S was asked to rate "to what extent he felt socially required to show respect to each of the 'professors.'" Thus, the higher pressures felt by S the higher should be his score on "predicted deference." The results of this analysis are shown in Table 4. Although there are very definite trends in the expected direction the differences just fail to reach an acceptable significance level.

This assumption also leads to the expectation that Ss who felt high pressure should in their behavior manifest higher deference than those who felt low pressure. This conjecture was fully supported by the data shown in Table 5. It is interesting, however, that the results seem more significant and the differences more pronounced with respect to observed than with respect to predicted deference. While the reasons for this are probably quite complex, a plausible one may be suggested. The scores of predicted deference were obtained from answers to scale items, while the scores of observed deference were obtained by observing Ss' behavior in a standardized particular situation. In this study, the scale was intended to measure a general response of

the Ss—the way the individuals generally behave. Their observed behavior, however, was in response to a very particular situation in which particular "professors" and particular task-orientations existed. The scores of "felt pressure" were obtained with reference to the particular experimental situation rather than with respect to the general case. It should, therefore, be expected that a higher correlation between felt situational pressure and observed behavior would be obtained.

SUMMARY

An experimental study was conducted in which normative conflict of foreign students was examined. From a treatment of norms as forcefields, predictions were made about the resolution of intergroup norm conflict under two conditions of situational potency of membership. It was hypothesized that the resolution of intergroup norm conflict favors the norm of the group whose situational potency is increased.

The results obtained did not show the clarity necessary for a definitive validation of the hypotheses. The major analyses led to acceptable levels of significance in only about half of the tests made; nevertheless, all the results presented a consistent pattern in harmony with the theoretical formulation.

REFERENCES

1. CHILD, I. L. *Italian or American? The second generation in conflict*. New Haven: Yale Univ. Press, 1943.
2. COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, 4, 512-532.
3. FRENCH, J. R. P., JR. A formal theory of social power. *Psychol. Rev.*, 1956, 63, 181-194.
4. LEWIN, K. *Resolving social conflicts*, G. W. Lewin (Ed.). New York: Harper, 1948.
5. LEWIN, K. *Field theory in social science*, D. Cartwright (Ed.). New York: Harper, 1951.
6. ROMMETVEIT, R. *Social norms and roles*. Minneapolis: Univ. of Minnesota Press, 1955.
7. SMITH, M. B. Cross-cultural education as a research area. *J. soc. Issues*, 1956, 12, (No. 1) 3-8.
8. SPINDLER, G., & GOLDSCHMIDT, W. Experimental design in the study of culture change. *Southwest. J. Anthropol.*, 1952, 8, 68-83.
9. STOFFER, S. A. An analysis of conflicting social norms. *Amer. soc. Rev.*, 1949, 14, 707-717.
10. VOGT, E. Z. Navaho veterans: a study of changing values. *Peabody Museum of Harvard Univ. Papers*, Vol. 41, No. 1.
11. ZAJONC, R. B. Aggressive attitudes of the "stranger" as a function of conformity pressures. *Hum. Relat.*, 1952, 5, 206-216.

⁸ Information regarding normative ideals was obtained by asking the S to indicate on the abscissa how much deference should students in America and India show to professors.

INDIVIDUAL AND SITUATIONAL DIFFERENCES IN DESIRE FOR CERTAINTY¹

ORVILLE G. BRIM, JR.

Russell Sage Foundation

AND

DAVID B. HOFF²

University of Wisconsin

HUMAN beings characteristically seek to understand their environment and to make it predictable. Explanations of this human trait have commonly involved motivational concepts such as needs for knowledge, certainty, security, structure, understanding, or cognition. Although the motivational terms listed have been used in similar ways, "certainty" seems most closely to represent the aspect of behavior measured by the major instrument in our studies.³ The purpose of the studies was to investigate the consistency of individual differences in strength of desire for certainty, and to see if this desire could be situationally increased or decreased by experimental procedures.

STUDY I: INDIVIDUAL DIFFERENCES IN DESIRE FOR CERTAINTY

Studies of the consistency of individual differences of the type considered here have not been numerous. Cohen *et al.* (6), in a study of need for cognition, report significant correlations between scores on different measures of the need. The measures used are not directly comparable to our own, yet they are similar in that they involve efforts to understand and structure the environment. Frenkel-Brunswik's well-known studies of intolerance of ambiguity (14, 15), and the study by Coulter (7) as summarized in Eysenck (12), also report individual consistency between different measures; again, the behavioral indices of intolerance of ambiguity appear similar to those used here.

We have also considered, tentatively, that studies of one type of response set (8) may

¹ The second study reported here is based on an unpublished master's thesis (17). The first study is based on research done at the University of Wisconsin supported by funds granted by the University Research Committee. A report of the first study was read at the annual meeting of the American Sociological Society in Urbana, Illinois, 1954.

² Currently with the United States Air Force.

³ In previous reports (4, 17) we tentatively had used the concept of need for security. Two recent reviews (5, 19) of use of the concept of security suggest that it is inappropriate in this context.

be relevant: the tendency to make extreme responses to test items. Studies of this tendency (2, 23) report consistent individual differences in extremity of response to various items within a test, and also consistency in total extremity scores from one test to another. If extremity of response may be viewed as an effort on the part of an individual to achieve a greater degree of structure and ordering of his world, the results parallel those of the studies mentioned above.

Our procedure was to obtain measures of individuals' desire for certainty, and to correlate the scores on these measures. In each of the substudies which follow two measures were used: a test constructed for this purpose, and the response-extremity score on a judgment or attitude test. The general hypothesis was that significant correlations would be found between scores on these measures.

METHOD

Measures

*Desire-for-certainty test.*⁴ The projective measure developed for these studies has been presented elsewhere (4) in detail. The test consisted of 32 statements about everyday events in the following form: "The chances that an American citizen will believe in God are about — in 100." The Ss filled in some probability value for each statement. In addition Ss indicated how sure they were of their estimates by rating the estimate on a five-point scale ranging from 1 (Very Sure) to 5 (Not Sure at All).

The statements included four from each of eight different areas, e.g., economics, recreation. Half of the statements were in conformity to generally accepted American values, while half were not. In addition, each statement had a different estimated probability value drawn from a set of 32 values ranging from 1 to 46 and from 54 to 99. Combinations of these three criteria

⁴ (a) The reliability of the test on an odd-even split, corrected by the Spearman-Brown formula, is .81 based on an *N* of 50 Ss; (b) the plot of over 500 test returns appears normally distributed, although no tests of normality have been made; (c) more than 200 test returns show no relationship between scores and standard socioeconomic variables; (d) a sample of 100 test returns shows no relationship (*r* equals $-.07$) with intelligence. With regard to the last two points especially, it should be noted that the sample consists of college students; homogeneity in pertinent respects may account for the lack of relationships.

were made randomly as specifications for the construction of each test statement.

Underlying present use of this test is the assumption that a strong desire for certainty is expressed in two ways: in a tendency to set probability values near the two extremes of 0 and 100, and in a tendency to claim a high degree of certainty that these values are correct. By jointly claiming a high certainty that something will definitely happen (or definitely not happen) uncertainty is reduced.⁵

The theoretical point of maximum uncertainty was taken on a priori grounds as probability equals .50, and "Not Sure at All" about this. Deviations in either direction from this point are regarded as an increase in certainty in scoring the test. It was scored by taking the distance of each probability estimate from its nearest end point, that is, 0 or 100, and multiplying this value by its associated certainty value, i.e., 1 for "Very Sure" through 5 for "Not Sure at All." The products are summed and averaged for the 32 items of the test. By this procedure the *lower* the score, the *higher* the desire for certainty.

Two other response determinants, in addition to effects of desire for certainty, were viewed as important. One was the amount of knowledge Ss had about an item. We attempted to control for effects of special knowledge by drawing the items from a variety of areas. However, there was no control over the effects of differences in general level of information, e.g., the college graduate certainly knows more in general than does the high school student. This defect would affect the results if groups of differing educational levels were to be studied. In the studies reported here, however, all Ss were college freshmen or sophomores so differences in general information should be minimal.

The second determinant considered important concerned whether or not the event in the test item was viewed as desirable. Probability estimates and other forms of judgment in ambiguous situations are distorted in a direction desirable to Ss (e.g., 9, 18). The selection of test items from different content and value areas was intended to randomize the effects of item desirability.

Attitude and judgment instruments. The specific instruments used are described for each substudy below. In all instances the instruments were scored for extremity of response. Two types of extremity were used as measures. One was the extremity of response to the *single* item, such as "How do you feel about television?" The other was the extremity of position on a *scale* of items for a single content area, such as one's favorable or unfavorable position on an attitude scale concerning China.

⁵ The use of both the probability estimate and its certainty in the test scoring is based on this point. Analysis shows that the joint score has in fact more validity as a measure than either one separately. Correlations with the personality ratings of roommates, reported in this paper, are +.18 and +.24, respectively, for the separate measures, and +.39 for the joint scoring method. See (4) for further discussion and evidence on this point.

Subjects and Procedures

The Ss for all studies were college students enrolled in introductory sociology courses. Different Ss were used in each study. In all instances the attitude or judgment instrument preceded the desire-for-certainty test in administration at a regular class meeting. Further details are given for each substudy below.

RESULTS

Substudy A. Thirty-eight Ss were presented with six attitude statements from different areas, e.g., "The Republican Party is doing a good job of government for the country." The Ss indicated their agreement or disagreement on a 13-point continuum.

Since relation between responses to any *single* statement and certainty-test scores should be curvilinear, certainty being associated with both extremely positive and negative responses, *etas* is the appropriate measure of relationship. *Etas* were computed for each item and none was significant.

To broaden the base of measurement, an extremity score was computed for each S for the six attitude items combined, by determining his average distance from the midpoint of the 13-point continuum on each of the six items, without regard to direction of deviation. The appropriate measure of relation between this average extremity score and test score is *r*, since here the relation should be rectilinear. In this instance, *r* equaled +.57, *p* < .01, in accord with our hypothesis.

Substudy B. Forty-nine Ss rated their roommates on an 8-point continuum on each of 21 personality traits. The rating scale was that developed by Fiske (13), as presented in McClelland (20), omitting item 22 because the wording seemed too difficult for Ss.

As in Substudy A, *etas* were computed for the relation between responses to each *single* item and the test scores, and none proved to be significant.

Again following a procedure similar to that of Substudy A, average extremity scores, based on the distance in either direction from the midpoint of the 8-point continuum, were computed for each S to obtain data of greater stability. As the appropriate measure of the rectilinear relation of these scores to desire for certainty, *r* in this instance equaled +.39, *p* < .01, again according with the hypothesis.

Substudy C. One hundred ninety-three Ss were administered the standard 30-item form

of the F scale (1, p. 260). Here our concern is with extremity of scale position rather than of item response, and our general hypothesis proposed that high and low scorers on the F scale would be similar in showing a high desire for certainty on the certainty test. Using eta as a test of curvilinear relation, a value of .28 was obtained, which has a probability (two-tailed) between .05 and .10 both for significance and for nonlinearity. Moreover, the rectilinear relation between test scores and F-scale scores, which we would expect to be insignificant, had an r of only +.06.

This result, which conflicts with earlier reports of correlates of F-scale position, can be viewed as providing weak support for the hypothesis. Most important is the fact that intolerance of ambiguity, the behavioral indices of which are similar to our indices of desire for certainty, is reported (14, 15) to have a positive *linear* relation to authoritarianism, which is represented by a high score on the F scale. On the other hand our findings are in accord with those of more recent studies (10, 11) in which *no* reliable linear relation was found between F-scale scores and intolerance of ambiguity. Davids (10) has recently compared and discussed the conflicting results of these studies.

Substudy D. Eighty-three Ss were administered three Thurstone scales on censorship, patriotism, and the Bible.⁶ Our hypothesis was that the relation between position on the Thurstone scales and desire for certainty would be curvilinear. Etas were thus computed for each of the three scales but none was found to be significant.

Data of greater stability were sought by combining Ss' responses on all three of the scales. Since the scales have no arbitrary midpoint (as did the measures in Substudies A and B) from which average response distance can be determined for the three scales, we used a midpoint for each scale the mean score of the group on that scale. The difference from each mean for each S was determined, and the average of the three differences obtained. These average extremity scores on the scales should have a rectilinear relation to test scores,

and thus, again, the suitable measure is r . In this instance r equaled +.28 $p < .02$.

STUDY II: SITUATIONAL DIFFERENCES IN DESIRE FOR CERTAINTY

Apparently there has been little research on situationally induced changes in needs for understanding, order, and the like. We were unable to find any such study directly concerned with certainty. Smock's recent study (24) seems most closely related; his results suggest that intolerance of ambiguity increases after exposure to stress, which consisted of threats to Ss self-image.

In our study Ss were frustrated and satisfied in tasks involving achievement of perceptual order and clarity, and subsequently administered measures of desire for certainty. We hypothesized that frustrated and satisfied Ss would show respectively higher and lower strengths of desire for certainty on the subsequent measure. Measures of desire for certainty were also obtained from a control group after exposure to "normal" situations. Our expectation was that strength of desire for certainty for this group would fall midway between that found in the two experimental groups.

METHOD

Instruments. A group of Gottschaldt's hidden figures (16) was used to present Ss with an ambiguous situation and to elicit their efforts to achieve clarity and organization. This group, consisting of five pages of figures, was the same as that used by Pacht (21). On each page one figure, or one or the other of two figures, was contained in a series of figures. The Ss were required to trace the sample figures which were hidden in the series on the page. The first and second pages were relatively easy, but subsequent pages became increasingly difficult. The Ss were instructed that this was a test of ability and they should do as well as they could in finding the hidden figures. The same measure of desire for certainty was used as in Study I.

Subjects and Procedures

Frustration condition. The Ss consisted of 57 college students, including both males and females, who were enrolled in three discussion sections of an introductory sociology course. During a regular class meeting all five pages of the Gottschaldt figures were presented under limited time conditions. Upon termination of the allotted time, this booklet was collected and was immediately followed by the desire-for-certainty test.

Satisfaction condition. The Ss consisted of 69 persons with the same characteristics as the frustration group. Procedure was identical with that above except that only the first two (easy) pages of the Gottschaldt fig-

⁶ Permission for the reproduction of the Thurstone Attitude Scales was granted by the University of Chicago Press. The three scales are out of print, and Dr. Leonard D. Goodstein of the State University of Iowa very kindly lent copies to the authors.

TABLE 1

STRENGTH OF DESIRE FOR CERTAINTY UNDER THREE CONDITIONS*

Comparison of conditions	Difference between means	<i>t</i>	<i>p</i> (two-tailed test)
Frustration vs. Control	7.3	2.33	< .05
Satisfaction vs. Control	7.4	2.48	< .05
Frustration vs. Satisfaction	14.7	4.12	< .01

* Range of test scores was 25-149. Means were: Frustration, 67; Control, 74.3; Satisfaction, 81.7. Higher score indicates lower need strength.

ures were used, and sufficient time was given so that all but a few Ss completed the two pages.

Control condition. An effort was made to sample "normal" situations by using 102 male dormitory-resident college students. On a given evening the desire-for-certainty test was administered to all occupants who were present in the dormitory rooms. The assumption was that Ss procured in this manner would be experiencing a variety of situations, e.g., studying, reading, visiting, writing letters.

RESULTS

The results of the study are presented in Table 1, which indicates clear support for the hypotheses. The mean scores for the two experimental groups differ reliably from each other in the expected direction. The scores also differ reliably from the control group score in the expected direction.

DISCUSSION

The results of these two studies suggest that individual differences in desire for certainty are consistent from one situation to another, and that this desire can be increased or decreased experimentally. The results parallel those of other studies of behavior directed toward ordering or stabilizing the environment. Among the many further research problems to which these findings lead, one seems to us of strategic importance: the relation between scores on the measures used in the various studies, e.g., the measures of intolerance of ambiguity, of response set, of need for cognition, and of desire for certainty. At present we have no empirical basis for assuming that these various efforts are manifestations of the same underlying construct. Substantial correlations, however, would suggest the existence of a single underlying need. Further work might then be profitably devoted to extending our information (e.g., 3, 14, 15, 22) on per-

sonality correlates of individual differences in need(s) of this kind, and to the questions of origins and development of such individual differences.

SUMMARY

Efforts by individuals to understand their environment and make it predictable are frequently explained by motivational concepts, among them the desire for certainty. Two studies of the desire for certainty are reported. In one study, the correlation was determined between Ss' extremity of response on several attitude and judgment instruments and their scores on a test of desire for certainty. The latter test asked for Ss' estimates of the probability of various events and for the certainty of these estimates. Responses were scored for the tendency to make estimates approaching 0 or 100, and to claim high certainty for these. Significant correlations were obtained between response extremity and test scores, suggesting that individual differences in strength of desire for certainty are consistent from one measure to another. The second study compared two groups on strength of desire for certainty, using the same test as above following the exposure of one group to a series of difficult Gottschaldt hidden figures, and the other to an easy group. The test was also administered to a control group. Significant differences between frustrated, satisfied, and control groups in desire for certainty were obtained, suggesting that strength of this desire can be increased and decreased by such situational influences.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- BERG, I. A. The reliability of extreme position response sets in two tests. *J. Psychol.*, 1953, **36**, 3-9.
- BERG, I. A., & COLLIER, JOANNE S. Personality and group differences in extreme response sets. *Educ. psychol. Measmt.*, 1953, **13**, 164-169.
- BRIM, O. G., JR. Attitude content-intensity and probability expectations. *Amer. sociol. Rev.*, 1955, **20**, 68-76.
- CAMERON, W. B. & MCCORMICK, T. C. Concepts of security and insecurity. *Amer. J. Sociol.*, 1954, **59**, 556-564.
- COHEN, A., STOTLAND, E., & WOLFE, D. M. An experimental investigation of need for cognition. *J. abnorm. soc. Psychol.*, 1955, **51**, 291-297.
- COULTER, T. An experimental and statistical study of the relationship of prejudice and certain per-

- sonality variables. Unpublished doctor's dissertation, Univer. of London Library, 1953.
8. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1951, **11**, 16-22.
9. CRONBACH, L. J., & DAVIS, BETTY MAE. Belief and desire in wartime. *J. abnorm. soc. Psychol.*, 1944, **39**, 446-458.
10. DAVIDS, A. Some personality and intellectual correlates of intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415-420.
11. ERIKSEN, C. W., & EISENSTEIN, D. Personality rigidity and the Rorschach. *J. Pers.*, 1953, **21**, 386-391.
12. EYSENCK, H. J. *The psychology of politics*. London: Routledge & Kegan Paul, 1954.
13. FISKE, D. W. The consistency of the factorial structures of personality ratings from different sources. *J. abnorm. soc. Psychol.*, 1949, **44**, 329-344.
14. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108-143.
15. FRENKEL-BRUNSWIK, ELSE. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 356-419.
16. GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren: I. Über den Einfluss gehaufter Einprägung von Figuren auf ihre Sichtbarkeit in umfassenden Konfigurationen. *Psychol. Forsch.*, 1926, **8**, 261-317.
17. HOFF, D. B. Some conditions affecting the need for security. Unpublished master's thesis, Univer. of Wisconsin, 1955.
18. IRWIN, F. W. Stated expectations as functions of probability and desirability of outcomes. *J. Pers.*, 1953, **21**, 329-333.
19. KNUTSON, A. L. The concept of personal security. *J. soc. Psychol.*, 1954, **40**, 219-236.
20. MCCLELLAND, D. C. *Personality*. New York: William Sloan Associates, 1951.
21. PACT, A. R. Changes in size constancy as a function of rigidity and stress. Unpublished doctor's dissertation, Univer. of Wisconsin, 1953.
22. RUBIN-RABSON, GRACE. Correlates of the noncommittal test-item response. *J. clin. Psychol.*, 1954, **10**, 93-95.
23. RUNDQUIST, E. A. Response sets: a note on consistency in taking extreme positions. *Educ. psychol. Measmt.*, 1950, **10**, 97-99.
24. SMOCK, C. D. The influence of psychological stress on the "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1954, **50**, 177-182.

Received May 21, 1956.

MODIFICATION OF AN EMOTIONALLY HELD ATTITUDE THROUGH ROLE PLAYING¹

FRANCES M. CULBERTSON

Arlington, Virginia

DESPITE the wide use of role playing in therapy and training, and the existence of a voluminous literature on the subject, there are few experimental studies concerning its effects, and none dealing with its ability to change an emotionally loaded attitude. This study is intended to contribute to the filling of this gap.

It is generally agreed that attitudes toward minority groups are difficult to change. Welt-fish noted that: "It seems probable from evidence to date that many individuals must go through what amounts to a therapeutic experience rather than the more typical conception of an educational procedure before ego-anchored prejudices can be yielded up for new sources of satisfaction" (8). Lewin's concept of "social habit" also embodies the factor of resistance to change. He observed: "To overcome this inner resistance, an additional force seems to be required, a force sufficient to 'break the habit,' to 'unfreeze' the custom" (4). This experiment attempted to explore the reversal of roles in role playing as such an influence on one kind of "social habit," the attitude of white people toward the Negro.

Psychodramatic literature indicates that the dynamic elements that differentiate role playing through role reversal from other learning situations are three: (a) role reversal, the taking on of attitudes, feelings, and behaviors that differ from one's own; (b) projection of needs and feelings, Moreno's "s" factor (6); and (c) development of self-insight. These dynamic factors should affect the experiencing individual's life space, including his attitudes. However, the potency of these forces for shifting "customs" would seem to depend upon the specific nature of the psychodramatic experience. Pearl Rosenberg (7) and Janis

and King (3) have called attention to one variable that may be significant in this respect, the element of role position. Their findings indicate that the closer a person is to a role—for example, as a participant rather than an observer—the more likely are his cognitive and motivational dispositions to be affected by the experience.

Another variable that may limit and modify the effectiveness of a psychodramatic session in changing attitudes toward a minority group is the "authoritarian personality" syndrome as measured by the F scale (1). "Authoritarianism" so defined should be related both to role-playing ability and to attitudes toward the Negro. Individuals high in authoritarianism should be relatively unable to expose themselves to the unstructured experiences of psychodrama, which jeopardize their defenses against repressed impulses; their minority group prejudice should therefore be relatively little affected by the psychodramatic experience.

Three hypotheses were therefore formulated for investigation: (a) A role-playing experience involving favorable attitudes toward integration of Negro and white housing results in favorable modification of the pertinent specific attitude, and to a lesser degree, in a more favorable generalized attitude toward the Negro; (b) the degree of favorable change resulting from the psychodrama is greater among participants than among observers; and (c) the degree of favorable change in attitude toward housing integration is greater among low "authoritarian" Ss than among high "authoritarian" Ss.

METHOD

To test these hypotheses the traditional three-stage design was used: (a) a pretest of Ss' specific and generalized attitudes toward the Negro, (b) a psychodramatic procedure directed at changing attitudes, and (c) a posttest of the specific and generalized attitudes. A control group was employed.

Subjects. Ninety-five students, 20 in an evening course on marriage and the family, and 75 in three separate classes in introductory sociology at a junior college in an east central state, were the experimental

¹ This paper is based upon a thesis submitted to the Department of Psychology of the University of Michigan in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to express her appreciation to Dr. Daniel Katz for his generous advice and assistance in connection with the study.

Ss. Twenty students in an introductory psychology course, who participated only in the pretest and posttest, were the control Ss. All of the data were gathered during the spring semester of 1954.

Attitude measures. The first measurements were made during a regular class period by the instructor of the class. The Ss were told that the purpose of the scales was to establish norms for some groups in the area. The specific attitude measured was that toward allowing Negroes to move into neighborhoods that had previously been occupied only by whites. The six alternatives presented for choice ran from "I am completely opposed . . ." to "I am completely in favor of allowing Negroes to move into white neighborhoods." Each S was required to rank the statements in the order in which they represented the person's own choices. His attitude position was obtained by coding his first three rankings. Using only transitive rankings, the three ranks gave a 16-point scale. For example, on the pretest, with statement *a* the most favorable, *b* the next, etc., the ranking *abc* was assigned position 1, the most favorable attitude toward Negro integration in housing; *bac* was assigned position 2, and so on to position 16, *fed*. The discrepancy in scale position between pretest and posttest yielded an index of change in the specific attitude. Posttest alternatives were presented in the opposite order from pretest ones.

The measure of change in generalized attitude toward the Negro consisted of the discrepancy between total pretest and posttest scores on Likert's scale (5). Since question 10 on this scale concerns attitude toward Negro integration in housing, Ss whose change in total score was due to this item alone were considered as not having changed.

The Ss at the favorable extreme on the specific attitude and on the Likert scale were excluded from the analysis, since they could not show change in the direction of the roles played or observed. There remained for analysis 87 Ss with respect to the specific attitude, of whom 45 were role players and 42 were observers, and 90 for the Likert scale, of whom 46 were role players and 44 role observers. Code numbers were used throughout the study to maintain Ss' anonymity.

Role-Playing Procedure. The role playing was conducted by *E* two weeks after the pretest measures had been obtained, in the guise of a laboratory session on human relations. The problem of the psychodrama was this: A considerable increase in the number of Negroes in the community was expected because of movement into the area of a defense plant. It was agreed that there would not be segregation, as there had not been any in the past. There was, however, concern about the possible development of tensions and antagonisms during the adjustment period. To minimize this danger, an educational program was planned by the local government. What should be the major theme of this educational program? Each of the three roles of the psychodrama involved the presentation of and advocacy of a specific theme for the educational program. Each of the themes was completely in favor of integration of the Negroes.

The Ss were divided randomly into groups of six, each group consisting of three participants and three observers, with each observer instructed to associate himself with an assigned role player. Members of each

group were arranged in a circle. In order to facilitate group discussion, the role players were instructed to arrange themselves in alphabetical order according to the names of the roles they were to play. This arrangement permitted them to call one another by name, and also permitted each observer to remember who had the particular role with which he had been instructed to associate. Before the role playing began, all Ss were given three minutes to consider how they would act the given role. At the end of this time, the 15 minutes of role playing began.

The posttest measures were administered seven to ten days later as a class exercise. A number of questions were added to determine whether Ss knew the purpose of the study, and whether their ideas on this had influenced their responses.

RESULTS AND DISCUSSION

Analysis of the data was in terms of chi-square comparisons of frequencies of positive change. Since the study centered on directional hypotheses a one-tailed test was used in chi-square tests for one degree of freedom. Differences at the .05 level and below were accepted as significant. The results of the study are summarized in Table 1.

The first objective was to test whether a psychodramatic experience can change an emotionally held attitude. In regard to the specific attitude toward Negro integration in housing, a chi-square value of 18.2 ($p < .0005$) was obtained in the comparison between role players and control group, and a chi-square value of 7.2 ($p < .0045$) for role observers and control group. In regard to general attitude toward the Negro, corresponding chi squares were obtained of 19.3 ($p < .0005$) for the com-

TABLE 1
FREQUENCY OF POSITIVE CHANGES

	N	Number Showing Positive Change	Comparisons, χ^2 and p values
Integration in Housing			
Control Group	18	2	χ^2 18.2
Role Players	45	30	$p < .0005$
Control Group	18	2	χ^2 7.2
Role Observers	42	18	$p < .0045$
Role Players	45	30	χ^2 5.99
Role Observers	42	18	$p < .008$
F Scale—above median	48	20	χ^2 9.16
F Scale—below median	39	28	$p < .0023$
Generalized Attitude			
Control Group	19	4	χ^2 19.3
Role Players	46	35	$p < .0005$
Control Group	19	4	χ^2 8.3
Role Observers	44	25	$p < .0035$
Role Players	46	35	χ^2 4.67
Role Observers	44	25	$p < .018$

parison of role players and control group, and 8.3 ($p < .0035$) for the comparison of role observers and control group. Changes of the sort hypothesized were clearly produced by the psychodrama, and persisted 7-10 days until the time of the posttest.

The second aim was to investigate whether positions in a psychodrama as participant and as observer result in different degrees of attitude change. The findings are positive and significant here also. Participants to a greater degree than observers shifted in the direction of the role experience. They changed more in the specific attitude toward Negro integration in housing, as indicated by a chi square of 5.99 ($p < .008$), as well as in attitude toward the Negro in general, as indicated by the chi-square value of 4.67 ($p < .018$) on the Likert scale.

The third objective was to explore the relationship between "authoritarianism" and attitude change through psychodrama. The data show that high-F Ss, those above the median, are significantly less likely to be influenced by role playing in their attitude toward housing integration than low-F Ss, those below the median, chi square being 9.16 ($p < .0023$). A next step suggested by these findings would be to investigate with resistant Ss the separate and combined effectiveness of role playing and role observing.

Differences in motivation and involvement.

The confirmation of the hypotheses underlying this study leaves open the nature of the mediating processes induced by the psychodrama. The difference in results between role players and role observers indicates—since both had the same information—that factors other than information were influential in reducing prejudice. Also, since high-F Ss would be expected to exhibit high degrees of conformity, the fact that low-F Ss changed in attitude to a greater degree than high-F Ss suggests that the diminishing of Negro prejudice was not merely a result of conformity pressures in the experimental situation. To what, then, can the change be attributed?

The theory underlying this investigation assumed that projection of feelings and self-insight were modifying forces. Supplementary data obtained from a questionnaire given immediately after the psychodrama showed that role players were more highly self-involved

in the drama than observers, as evidenced by their emotions, perceptions, and judgments regarding the role experience. The following factors were all present to a greater degree—significant beyond the .05 level—for participants than for observers: (a) degree of association with assigned role, (b) time during which attention was paid to the given role, (c) emotional involvement with the role, (d) perception of the drama in terms of feelings, thoughts and emotions rather than in terms of the situation and actions of the characters, (e) positive rather than negative solutions regarding the role problem, and (f) noncritical rather than critical judgment regarding the role behavior and role session. Thus role participation appears to have led to greater degrees of motivation and involvement in the drama than role observation. The part played by projection and self-insight in these effects, however, remains to be explored.

SUMMARY

This experiment investigated (a) whether role playing can change an emotionally loaded attitude, (b) whether a participant position in a role-playing session is more likely to result in a change in attitude than an observer position, and (c) whether Ss high in "authoritarianism" are less affected by a psychodramatic experience than Ss low in "authoritarianism."

Ninety-five experimental Ss responded to questionnaires before and after a role-playing experience: (a) a measure of attitude toward "allowing Negroes to move into White neighborhoods," the specific attitude associated with the role session, and (b) the Likert scale measuring the generalized attitude toward the Negro. Twenty control Ss took similar measures, without exposure to the psychodrama. The attitude change procedure was a 15-minute role-playing session in which Ss were randomly assigned either to be participants or observers with particular roles. All roles were completely in favor of Negro integration in housing.

The findings were significant and in the predicted direction. Experimental Ss shifted their attitude toward Negro integration in housing as well as their general attitude toward the Negro in the direction of the role experience. Role players shifted to a greater degree than

observers in a favorable direction. Those Ss below the median on the F scale changed more in a favorable direction in attitude toward Negro integration in housing than those above the median on the F scale. Findings on degree of self-involvement indicate that role players were more motivated and more involved in the drama experience than role observers.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. CULBERTSON, FRANCES M. The modification of emotionally held attitudes through role playing. Unpublished doctor's dissertation, Univer. of Michigan, 1955.
3. JANIS, I., & KING, B. T. The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **2**, 211-218.
4. LEWIN, K. Group decision and social change. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 330-344.
5. LIKERT, R. A technique for the measurement of attitudes. *Arch. Psychol.*, 1932, No. 140.
6. MORENO, J. L. *Who shall survive*. New York: Beacon House, 1953.
7. ROSENBERG, PEARL P. An experimental analysis of psychodrama. Unpublished doctor's dissertation, Harvard Univer., 1951.
8. WELTFISH, GENE. Racial and religious prejudices in everyday living. *J. soc. Issues*, 1945, **1**, No. 1, 51.

Received May 25, 1956.

ANXIETY AND VERBAL LEARNING¹

SHELDON J. KORCHIN² AND SEYMOUR LEVINE

The Institute for Psychosomatic and Psychiatric Research and Training, Michael Reese Hospital, Chicago

THERE is little question that anxiety, particularly in the intensities seen in psychiatric patients, can have distinctly disruptive effects on learning as well as on other types of cognitive behavior (for reviews, see 2, 5, 9). At the same time anxiety can organize and facilitate behavior, not only where the measured performance is directly related to anxiety-reduction, but over a wider range of behavior by raising the general level of activity or vigilance. Whether anxiety facilitates or disorganizes is probably a function of the quality and intensity of the anxiety (2), the relevance of the performance for anxiety-reduction, and the prior personality organization of the subject.

In learning, the effects of anxiety vary with the type of material to be learned. Considerable data on this point have been accumulated in the recent attempt by Taylor, Spence, and their co-workers to conceptualize the anxiety-learning problem within the context of Hull's learning theory (16, 18). Experiments have shown that anxious Ss perform less well in maze learning (4, 12), serial rote learning (10, 14, 18), in general in more difficult and complex learning tasks. On the other hand, anxious Ss show more rapid conditioning to noxious stimuli (3, 16, 19).

In this experiment Ss learned two lists of paired-associate items, adapted from an early experiment of Ruch's (15), which differed sharply in their familiarity, difficulty, and most important in their relation to pre-existing habits. The first of these consists of familiar words in logically-related pairs; the second of incorrect multiplication formulas, e.g., $2 \times 3 = 9$. In the first case learning is congruent with greatly overlearned habits, and in the second

it is in conflict with them. While there might be little difference in the learning of the word-pairs, it might be predicted that anxious Ss will perform proportionately less well with the false equations.

Many of the recent studies of anxiety and learning have used paper-and-pencil scales, most commonly that devised by Taylor (17), to distinguish criterion groups of more and less anxious Ss, usually among university students. In the present study we have used the same criterion and similar groups. This was done in part for convenience and in part to make the present findings comparable to a now considerable number of studies. In addition to high and low groups defined by the Taylor Manifest Anxiety Scale, we have also studied hospitalized psychiatric patients whose present behavior and history show considerable evidence of anxiety. Unlike even the most anxious students, these Ss are sufficiently disturbed so as to require care in a residential treatment center. The three groups of subjects—high and low normals, Taylor-defined, and anxious patients—may be conceived as differing not only in the extent of anxiety but in the personality resources for adjustive behavior generally.

A secondary purpose of this experiment was to explore whether different intertrial intervals (i.e., massed vs. distributed practice) influence the relationship between anxiety and the learning of each type of material. Learning in various situations has been found to be facilitated by the spacing of trials (7, 13), although less profound effects are reported for paired-associate than for serial learning (6). Although the explanations for the advantages of distributed practice vary, a number of theories emphasize some form of differential forgetting of interferences built up during learning. The question here is whether spacing of trials will affect equally the learning of each type of material and the performance of each of the subject groups.

METHOD

Subjects. A high-anxiety group (HA, $N = 18$) and a low-anxiety group (LA, $N = 18$) were selected from a group of 110 university extension students on the basis

¹ This investigation was supported in part by the Medical Research and Development Board, Office of the Surgeon General, Department of the Army, under Contract No. DA-49-007-MD-469 and in part by the National Institute of Mental Health, Public Health Service, under research grant M-1051. We are grateful to Dr. Helen Heath and Mrs. Judith Kandler for help with the analyses of the data.

² Now at the Research Division, Columbus State Institute of Psychiatry, Ohio State University.

of extreme scores on the Taylor Manifest Anxiety Scale (17). Students in the low group had scores of 10 or below with a mean of 6.5. The high group were selected from those with scores of 20 and above and had a mean of 25.4. In age, the groups were respectively 26.7 years and 28.6 years. The over-all range was from 17 to 44 years.

A third group (PA) consisted of 17 Ss selected from the psychiatric patient population of the Institute for Psychosomatic and Psychiatric Research and Training of the Michael Reese Hospital. These patients were considered, on the basis of history and clinical examination, to be anxious or to have considerable "anxiety proneness." In diagnostic terms, ten were called "anxiety reaction," five "depressive reaction," and the remaining two "schizoid personality" and "emotionally unstable personality." All of the patients had been recently admitted, most within a week prior to testing and were in psychotherapy at the time. For six of the group this was a second admission to this hospital. On the Taylor Scale this group had a mean score of 28.8, with a standard deviation of 7.9. Interestingly, only two Ss fell below the score of 20 used to define the HA group, one with a score of 17 and the other 8. The mean age of the patient group was 41.0 years, ranging from 20 to 55.

In studies of aged and young Ss on the tasks used in this experiment, learning was also found to be related to intelligence (8, 15). To check on the comparability of the student and patient groups in general intellectual level, the Vocabulary Scale of the Wechsler-Bellevue was administered. Although the patients score somewhat lower, there is no significant difference between them and the students. However, because of the existing differences between the groups in age, education, and intelligence, despite considerable overlap, the findings of this study will be presented primarily in terms of the comparisons between the high- and low-anxiety student groups, and between experimental treatments within each of the three groups.

The learning tasks. Two lists of paired associates originally developed by Ruch (15) were administered in a form recently modified by Korchin and Basowitz (8): (a) *Word Associates*: pairs of logically related and familiar words, e.g., Man—Boy; (b) *False Equations*: a series of incorrect multiplications, e.g., $3 \times 5 = 6$. Each list consisted of 8 items. After the initial presentation of the list, S was given 6 test trials. To minimize serial learning the order of items was changed on each trial. The word-associates task was always administered first. At the beginning of each task a sample pair was presented and the task explained to S. On each of the 6 test trials, S was first shown the left-hand member of the pair alone and asked to anticipate the right-hand member, after which the complete pair was presented. The items were projected by a motor-driven 35 mm. strip film projector in a semidarkened room, each frame remaining in view for 4 seconds. The instructions used were essentially the same as Ruch's (15). The S was told to guess if uncertain.

Delay and Immediate conditions. For half of the Ss in each group the trials were separated by 8 seconds, the time required for advancing two blank frames in the projector (Immediate condition). The remaining Ss were given an additional half-minute during which

TABLE 1
DIFFERENCES BETWEEN EXPERIMENTAL GROUPS
IN WORD ASSOCIATES AND FALSE EQUATION
TOTAL SCORES

Group	Condi- tion	N	Mean Total Score	
			Word Asso- ciates	False Equa- tions
Low Anxiety students (LA)	D	9	44.78	24.67
	I	9	42.56	22.67
High Anxiety students (HA)	D	9	41.22	16.33
	I	9	40.00	17.00
Anxiety patients (PA)	D	8	39.25	16.12
	I	9	28.56	7.88

the motor of the projector was turned off (Delay condition). During this period these Ss were engaged in conversation to offset the possibility of rehearsal. There were approximately 2 to 3 minutes between the two lists for all Ss.

In the HA and LA groups 9 Ss were tested in the Immediate and 9 Ss in the Delay condition. In the PA group there were 8 in the Delay and 9 in the Immediate condition.

RESULTS

Differences between groups in total performance. In Table 1 are presented the mean total scores for each of the groups for the two learning tasks. Each S's total score is the number of correct anticipations of the 8 items summed over the 6 test trials; the maximum score thus being 48.

It is clear on inspection that all groups performed better on the word than on the equation material. In none of the groups was there any S who achieved a higher score on the false equations than on the word associates.

Comparing the LA and HA groups, the students with the lower anxiety level performed somewhat better on the word associates, but were distinctly superior in learning the false equations. The differences between the delay (D) and immediate (I) conditions were uniformly small. To evaluate these differences, an analysis of variance was done on the word-associate scores and another on the false-equations scores. For the word associates the LA — HA difference approaches significance ($F = 3.770$, with 1 and 32 *df*, $p < .10$) and the comparison of I and D is insignificant ($F = 1.198$). However, on the false equation task the LA group performed significantly better than the HA ($F = 7.099$, with 1 and 32 *df*, $p < .025$). Again the I — D difference is insignificant ($F = .064$).

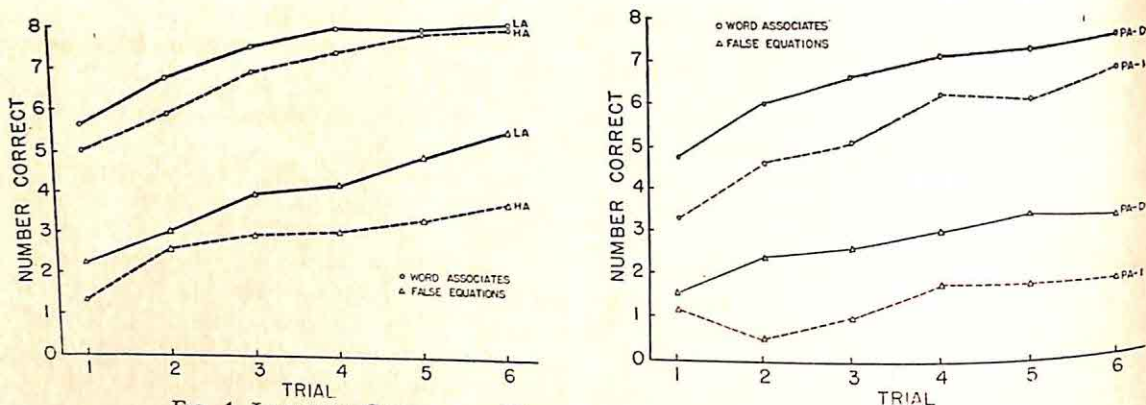


FIG. 1. LEARNING CURVES FOR WORD-ASSOCIATE AND FALSE EQUATION TASKS

LA and HA represent low- and high-anxiety normals, respectively, while PA-D and PA-I represent patient groups learning under "delay" and "immediate" conditions.

The patient group tested under delay conditions scored much like the HA; in the immediate condition, however, there was poorer performance in the PA-I group. For the word associates the difference between PA-I and PA-D approaches significance ($t = 1.84$, $p < .10$) and for the false equations it is significant at the $p < .05$ level ($t = 2.40$).

Rate of learning. The learning curves for the word-associate and false-equation tasks are presented in Fig. 1. Since there were no significant differences in total score between the I and D conditions within either the HA or LA groups, and on inspection the trial-by-trial curves are virtually identical, the Ss in the I and D subgroups have been combined for simpler presentation. For the PA group, however, the curves are separately drawn for those tested under immediate and those under delay conditions of presentation.

The LA Ss performed better than the HA on each trial on both tasks, although as already indicated in the comparison of total scores, the separation between the curves is greater in the false equations. Although the word-associate curves for the two student groups are essentially parallel, there does seem to be a difference in the forms of the functions between the HA and LA groups in the false-equation learning. The scores of the LA group increase consistently from trial to trial: by contrast after an initial first-to-second trial rise the HA group retains almost the same level through the remaining trials. As a simple measure of difference in slope, the difference between HA and LA in the amount of increase from second to last trial was tested by t test and found to be

significant at the $p < .02$ level ($t = 2.694$). Thus, on the equation material the more anxious Ss had relatively little learning after the initial trials, compared to less anxious Ss and as well to their own performance on the word associates.

Similarly, the patient groups have quite flat learning curves for the false equations, with the curve for the I group rising less, though not significantly so, than that for the D group.

Type of error made during learning. When the stimulus item first appears, S can either respond with the correct item or he can err in a number of distinct ways. The groups not only differ in the amount learned in each task, but closer analysis of the types of error made during learning shows further differences in learning behavior.

Three types of errors can be distinguished for both the word-associate and false-equation material: First, S may simply not respond, despite the instruction to guess if uncertain (*No response*). Second, S may offer a response word or number which exists nowhere in the list (*Random response*). Third, he may respond with an item which, although correct elsewhere in the list, is inappropriate for the presented stimulus (*Correct elsewhere*). For the false-equation items, S may in addition give a response which represents the correct multiplication or possibly addition or subtraction of the numbers in the stimulus equation. *Multiplication* errors are of particular theoretical interest since they may indicate the direct intrusion of a competing response tendency. Similarly, however, both *addition* and *subtraction* may also reflect habitual re-

TABLE 2
PERCENTAGES OF VARIOUS TYPES OF ERRORS MADE IN FALSE EQUATION LEARNING BY EACH OF THE EXPERIMENTAL GROUPS

Group	Condition	N	Total Number of errors	Percentage of Total Errors which are:					
				No Response	Random Response	Correct Elsewhere	Multiplication	Addition	Subtraction
Low anxiety students (LA)	D	9	210	19	13	44	2	18	4
	I	9	228	28	8	46	1	10	7
High anxiety students (HA)	D	9	284	25	5	52	3	10	5
	I	9	279	56	6	28	1	6	3
Anxiety patients (PA)	D	8	263	58	6	24	1	9	2
	I	7*	302	73	4	17	2	4	0

*For 2 Ss in this group the type of error made was not recorded. In addition, two Ss blocked completely on their 5th trial and were unable to continue. The total errors, and the percentages, are for the Ss and trials available.

sponses to the implied structure of an arithmetical problem. However, scoring these errors depends on interpretation, since the response might be "random" and in particular cases *Correct elsewhere*.

On the word-associate task, where there are fewer errors generally, occurring mainly in the early trials, the majority of errors are *No response* and the pattern does not differ significantly among the several groups.

More interesting is the analysis of the more difficult false-equation task. The distribution of errors for each of the groups and conditions is given in Table 2. The entries are percentages of the total number of errors made by each group on all six trials combined. For statistical evaluation, *U* tests were run on the arrays of individual *No response* proportions. Since the errors group into two major types, those in which *S* simply does not respond and those in which some sort of response is offered, the single value serves as an index to the patterning of responding-not responding behavior.

In general, there were relatively few of the "arithmetical" errors, whether multiplication, addition, or subtraction. Most of the errors were either *No response* or *Correct elsewhere*, and the balance between responding and not responding behavior differs importantly between groups. The LA group, with no significant difference between I and D conditions, had the greatest proportion of its errors of the *Correct elsewhere* sort. By contrast, within the HA group those tested under the I condition had a significantly higher proportion of *No response* errors ($p < .025$). Both patient subgroups had more *No response* errors than the student groups, and here again Ss in the immediate condition were more prone not to

respond, although the difference between PA-I and PA-D is short of statistical significance. Over all, it does seem that with the equation material there is an interactive effect of anxiety and the immediate condition of testing: at higher anxiety levels, the massed sequence of trials leads to an increased tendency not to respond rather than to offer some response to the stimulus items.

DISCUSSION

The performance of the three groups under the various conditions of testing can best be understood as a function of the *effective anxiety level* in the particular learning situation. The Ss vary in their initial anxiety and their ability to function under stress; the two learning tasks and the two conditions of intertrial spacing are conceived as differing in their stress value. The amount and rate of learning, and the types of errors made, depend on *S's* effective anxiety defined by his initial stress tolerance and the combined values of the stresses imposed on him.

Anxiety and the nature of the task. The expectation that the nature of the material to be learned is a decisive factor in the learning behavior of anxious Ss has been verified. More anxious Ss differ relatively little in the amount learned when dealing with simple and logically associated material. However, the groups are significantly distinguished in the learning of unfamiliar and difficult material contradictory to previous knowledge. We believe that this task constituted a greater stress for the anxious Ss.

The anxious individual can be conceived as generally less capable of integrative behavior. To perform well in any learning task, more

generally in any intellectual test situation, an individual must be able to accept the requirements of the task, segregate the relevant and important from the trivial and peripheral, establish and maintain task-appropriate sets, and modify these as the task requires. The optimal cognitive organization required for adjustive and ordered behavior is changed under anxiety either in the one direction of more diffuse and disorganized behavior or to the other extreme of greater rigidity. One consequence of this lessened integrative capacity results in the appearance in the test situation of what have variously been called "task-irrelevant responses" (11) or "anxiety-produced interferences" (10).

The more novel or difficult the task confronting *S*, the greater is the demand on the individual's ability and, simultaneously, the greater is the potential threat. Precisely the situation which requires most efficiency is most likely to evoke further anxiety in the already anxious *S*. This in turn acts to reduce performance further. In the situation in which *S* has to make a novel adjustment and can not utilize existing behavior patterns, the possibility of failure and the consequent loss of self-esteem can further release anxiety and further reduce *S*'s ability to develop appropriate behavior.

Three types of behavior may be distinguished in the anxious *S* under stress, all of which, however, have the common property of reducing effectiveness in a test situation. Either *S* exhibits defensive behaviors designed to reduce anxiety but which at the same time function as interferences in the present task, or he falls back on pre-existing methods of dealing with similar problems which are, however, inappropriate to the present situation. Either the defensive act or the habitual behavior may function as "anxiety-produced interferences," although in the latter case the resulting performance may more likely appear as "rigidity." Lastly, behavior in the threatening situation may become more completely disorganized and primitive and reflect simply the lack of integration rather than the interference of either defensive behaviors or irrelevant existing problem behaviors.

The immediate and delay conditions or intertrial spacing may be considered, as have the word and equation tasks, to differ in their stress value. There is little evidence of any

general effects on all *Ss* such as have been found in studies of massed and distributed practice and which might be understandable in terms of associational factors. Whether trials are spaced or not makes little difference in the learning of the least anxious group (LA). But with increased anxiety the massing of trials seems to act as an experimental stress—comparable to distraction, rapid pacing, and the like—and the extent of its effect is proportional to *S*'s stress tolerance. For the most anxious *Ss* (PA) even the additional 30 seconds between trials acts to reduce effective anxiety and results in improved performance.

Anxiety and response inhibition. Analysis of the types of error made during learning reveals important differences between groups in the relative frequency with which they offer or do not offer incorrect responses. As anxiety is increased, in this experiment through the interaction of initial anxiety level and the imposed stress of the experimental conditions, *Ss* are less capable or inclined to venture incorrect responses. This behavior, which may be called "response inhibition," was found in more extreme form in the comparison of aged and young *Ss* on the same learning tasks (8).

One can only speculate as to the meaning of response inhibition in anxiety. There are two basic assumptions from which interpretations of this behavior can be derived: (a) Production or verbalization of response is blocked, although there is no qualitative difference in what is learned; or (b) the learning process does differ in heightened anxiety states so that *S* acquires fewer items that can be offered as "correct elsewhere" responses, and hence does not respond. Clearly, these assumptions are not exclusive, for one could conceive both of differences in the mode of acquisition of response and simultaneously of blockage in response production as parallel processes.

On the first of these assumptions, it may be that with increased anxiety there is a greater "cautiousness" in the interest of ego defense. Although encouraged by the instructions to guess if uncertain, the anxious individual may require greater certainty before committing himself. Although he scores no better with "no response" than with an incorrect response, the inhibiting of response in an uncertain and threatening situation may reflect a more general ego-defensive constriction of behavior

through which inadequacy is denied and further anxiety averted.

Along with the defensive behavior, and perhaps because of it, more time may be needed at heightened levels of anxiety for the integration of response. The experimental procedure requires that *S* read the stimulus item and select and report a response in the four seconds during which the stimulus item remains in view. The anxious individual, less able to maintain integrative behavior generally and being more prone to interfering response tendencies, may require more time for the complete process of perceiving - judging - recalling - responding than does the less anxious individual.

Such interpretations assume that fewer incorrect responses of any sort are offered because of a blocking resulting from defensive needs, the interference of other response tendencies, or inability to survey and report a response in the available time. One need not assume that there is any qualitative difference in what is learned. However, recalling that the increased "no response" behavior is paralleled by a compensatory decline in the "correct elsewhere" errors, there may also be a somewhat different mode of learning in anxiety so that fewer such responses are acquired.

Normally, learning in a task of the sort employed here can be assumed to consist first of the identification and recognition of the response items in the list and thereafter of the establishment of specific stimulus-response pairings. One learns that 8, 7, and 1 belong in the list and are proper associates to *some* stimulus item and then increasingly differentiates the particular pairings—that 8 goes with 2×5 , and 7 with 5×1 , and so on. Errors made during learning would then consist of offering items which are "correct elsewhere" before the correct combination is learned. Giving "no response" may indicate less learning of the separate response items initially. With greater anxiety learning may consist principally of the formation of *discrete* stimulus-response pairings, and may to some extent bypass the earlier stage in which response items are learned as belonging *somewhere* in the series. This tendency to isolate the stimulus-response units within the learning task may be a particular aspect of the generally more constricted cognitive functioning found in anxiety in other types of situations.

Although the phenomenon of response in-

hibition is an important aspect of anxiety behavior, and has been found in this experiment to differ meaningfully between subject groups and conditions, it is clear that there is not sufficient basis here for judging among alternate lines of interpretation. The phenomenon should be studied further, perhaps by varying the instructional set, the time of presentation, or both, and by using other types of verbal learning material.

Anxiety and the rate of learning. Ausubel, Shiff, and Goldman (1) have shown that anxious *Ss* perform proportionately worse than less anxious *Ss* on the early trials of the learning tasks used in their experiments and, correspondingly, show greater improvement with subsequent practice. They argue that the greatest deficit should be shown in the novel and threatening situation and that the greater improvement of the anxious *Ss* is due at least partly to the "release from the fear-inspired, nonimprovising response set as increased familiarity with the learning task renders it less novel" (1, p. 545). By contrast, the more anxious *Ss* in our study show proportionately *less* practice effect in the learning of the false equations. The learning curves for the more anxious *Ss* tend to flatten and show relatively less improvement after the initial trials, quite the reverse of the greater acceleration reported by Ausubel, Shiff, and Goldman.

These findings are not at all contradictory, for the differences may be interpreted in terms of the stress-value of the tasks in the two experiments. Whether increased familiarity leads to subsequent improvement or not obviously depends on what happens in the initial experience with the task. The anxious *S* may be presumed to approach any new situation apprehensive as to the unknown danger and fearful that he may not be able to cope with it. If by failure and frustration in the initial trials these apprehensions are *confirmed*, then the threat is increased, anxiety raised, and subsequent performance blocked. Familiarity normally has the advantage of reducing potential threat except when, as seems to have occurred in this experiment, the experience itself provokes further anxiety.

SUMMARY

Three groups of *Ss*, differing in anxiety, learned two different types of paired-associate materials—a list of simple, logically-related

word-pairs and another of incorrect multiplication formulas. The Ss were two groups of university students selected from extreme scorers on the Taylor Manifest Anxiety Scale and a group of hospitalized anxious patients. Half of the Ss in each group were tested under "immediate" and half under "delay" conditions of intertrial spacing.

Among the students, the high and low groups, tested under both delay and immediate conditions, showed no significant differences in their learning of the word-associate material. With the more novel and difficult false equations, however, the high anxiety Ss learned significantly less over all and showed a somewhat slower rate of learning from trial to trial. Among the patients, those tested under the delay condition performed about as well as the high-anxiety students, while those under the immediate condition learned a lesser amount both of the word and equation materials.

An interesting aspect of anxiety behavior emerged in the study of the type of errors made during learning. Though not differing in word associates, in the false equation learning the degree of "response inhibition"—not responding rather than offering some response to the stimulus item—does vary with anxiety and intertrial interval. The least anxious Ss showed the least response inhibition and no difference between delay and immediate testing; the high-anxiety students were significantly higher in the immediate condition; and the patients with the highest level of response inhibition also rose, though not significantly, in the immediate testing. Possible interpretations of response inhibition are discussed.

The anxiety groups differ in their ability to tolerate stress, and the two tasks and the two conditions of intertrial spacing differ in their value as stresses. Performance, both in the amount and rate of learning and in the pattern of errors, is a function of S's effective anxiety level in the particular situation, which is defined by his initial anxiety and the combined value of the imposed stresses. The lower the initial anxiety, the less is the effect of differences in the experimental conditions; for more anxious Ss even small changes (e.g. massing trials) leads to sizable changes in performance. Some of the mechanisms through which heightened anxiety affects performance are discussed.

REFERENCES

1. AUSUBEL, D. P., SCHIFF, H. M., & GOLDMAN, M. Qualitative characteristics in the learning process associated with anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 537-547.
2. BASOWITZ, H., PERSKY, H., KORCHIN, S. J., & GRINKER, R. R. *Anxiety and stress*. New York: Blakiston Division, McGraw-Hill, 1955.
3. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615-623.
4. DIETHELM, O., & JONES, M. R. Influence of anxiety on attention, learning, retention and thinking. *Arch. Neurol. Psychiat.*, 1947, **58**, 324-336.
5. HANFMANN, EUGENIA. Psychological approaches to the study of anxiety. In P. H. Hoch & J. Zubin (Eds.) *Anxiety*. New York: Grune & Stratton, 1950. Pp. 51-69.
6. HOVLAND, C. I. Experimental studies in rote-learning theory. V. Comparison of distribution of practice in serial and paired associate learning. *J. exp. Psychol.*, 1939, **25**, 622-633.
7. HOVLAND, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 613-689.
8. KORCHIN, S. J., & BASOWITZ, H. Age differences in verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 63-68.
9. LAZARUS, R. S., DEESE, J., & OSLER, SONIA. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **4**, 293-317.
10. MALMO, R. B., & AMSEL, A. Anxiety-produced interference in serial rote learning with observations on rote learning after partial frontal lobectomy. *J. exp. Psychol.*, 1948, **38**, 440-454.
11. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166-173.
12. MATARAZZO, J. D., ULETT, G. A., & SASLOW, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, 1955, **53**, 79-95.
13. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
14. MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91-96.
15. RUCH, F. L. The differentiative effects of age upon human learning. *J. gen. Psychol.*, 1934, **11**, 261-286.
16. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-89.
17. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
18. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.
19. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83-91.

Received May 31, 1956.

RELATIONSHIP OF ASCENDANT-SUBMISSIVE BEHAVIOR IN DYADIC GROUPS OF HUMAN SUBJECTS TO THEIR EMOTIONAL RESPONSIVENESS¹

VLADIMIR CERVIN²

University of Toronto

IN human interaction the behavior of some individuals is often regarded as more important than that of others because it seems to have more weight in determining the character of their interrelationship. Such interrelationship may be called dominance, or more exactly, ascendancy-submission. The problem here is how the development of ascendancy-submissive behavior could be predicted. This paper describes the experimental investigation of one aspect of this problem, viz., the role of a personality variable (emotional responsiveness) in the development of an ascendancy-submissive relationship in pairs of subjects (Ss) having a free choice of verbal behavior within certain limits, and interacting over a relatively short period of time. The interaction consisted of discussion by two Ss of a topic on which they held opposite opinions at the beginning of the experiment. The personality variable was operationally defined in terms of the difference between the scores of the two Ss on a specially constructed Guttman scale of emotional responsiveness. Ascendant behavior of one S in a pair over the other was defined in terms of differences between their behavior measures, viz., greater speed (inverse of latency) of the first statement, larger amount of participation in the discussion, and greater resistance to change of opinion. A model of interaction between two individuals was constructed on the basis of stimulus-response behavior theory (3, 5), developed mainly with respect to one individual, and was used as a basis for derivation of predictions. In order to apply the model, the difference in emotional responsiveness was interpreted as a function of the difference in drive level (2, 7), and differences in behavioral measures as functions of differences in reaction

potentials (5). The model is a statement of the relationship between these two variables in time. From this theoretical relationship it was predicted that Ss with high emotional responsiveness would tend toward an ascendant position, and those with low emotional responsiveness toward a submissive position. Derivations of predictions are developed first in verbal, then in mathematical form, for psychologists interested in mathematical models.

A behavior theorem particularly useful for our purpose states that when two (or more) incompatible response tendencies occur in an organism simultaneously, the one with the highest reaction potential is evoked (5, p. 104). Generalized to the interaction of two individuals, this theorem predicts that the individual with the higher reaction potential should be in the ascendant position on all the three measures of behavior. Reaction potential is regarded as an increasing function of drive level, if other variables are kept equal. Drive level is an increasing function of the score of emotional responsiveness. The amount of an S's participation in the discussion, and the speed of his first statement, are direct functions of his reaction potential. Consequently, Ss with high emotional responsiveness scores should participate more in interaction and make their first statements earlier than Ss with low scores. Furthermore, interaction is interpreted as a learning process, with mutual reinforcement or extinction of Ss' response tendencies resulting from it. The amount of reinforcement and extinction received by each S depends on the behavior of the other S, and consequently (a) on his emotional responsiveness score (*ceteris paribus*), and (b) on the opinions supported by his responses. Hence, knowing Ss' learning functions, and certain other conditions, change of opinion can be predicted from knowledge of Ss' emotional responsiveness scores and initial opinions: more Ss with low emotional responsiveness should change their opinions to agree with Ss with high emotional responsiveness than vice versa.

¹ Research done under National Health Grant from the Government of Canada to Professor J. D. Ketchum.

² The writer is indebted to Professor Ketchum for advice and cooperation in carrying out this study. Thanks are also due to Dr. A. H. Shephard, Dr. E. Tulving, Mr. M. Humphries, and Mr. G. Groen, for criticizing parts of the manuscript, and to Joan Cervin for editorial help and typing.

The predictions of latency and the amount of participation are quite straightforward, but it seems necessary to give an explanation of the predictions for change of opinion. The *S*'s adherence to an opinion can be measured by the difference between the frequency of statements supporting it and that of statements supporting the opposite opinion. If differences between *S*'s competing reaction potentials are not too great, each *S*, because of behavioral oscillation, may not only make statements supporting his original opinion, but also some supporting the competing opinion. The frequencies of these two classes of statements should be in proportion to the difference in their respective reaction potentials. At the beginning of the experiment, the reaction potential of an *S*'s original opinion should be dominant, and he should support his own opinion more frequently. But both the dominant and the competing reaction potentials should change as a result of interaction, and so should the frequencies of responses supporting each opinion. Extinction of a response takes longer under high than under low drive, and the high-drive *S* gets less frequent reinforcement from his low-drive opponent (because of lower rate of response). Therefore, learning of a response which competes with a stronger one should take longer for high-drive *S*s, who thus should, within a specified time limit, change their opinions less frequently than low-drive *S*s. This position is analogous to that developed by Spence *et al.* for one individual (4, 7), but is not in line with a study by Buchwald and Yamaguchi (1).

To test the predictions, thirty pairs of *S*s with different, and thirty with equal, drive levels, were each given a topic to discuss. The *S*s' opinions on a series of topics of some interest to them had been ascertained before the experiment. For each pair a topic was selected on which its members held opposite opinions. Pairs were interacting under the chairmanship of a third *S*, who remained neutral. All pairs were instructed to arrive at an agreement if possible, or else to agree to disagree. The *S*s within each pair were equated for measures of rigidity (interference of irrelevant responses), their own ratings of the importance of the issue discussed, their initial opinions, and several personality characteristics. These precautions were taken in order to insure

that "other variables" influencing the reaction potentials should be as nearly equal as possible within and between pairs. Certain plausible assumptions about the interaction process, such as constant total rate of response (2), were also made.

METHOD

Scales

Two Guttman scales were constructed: the *E* scale of emotional responsiveness and the *R* scale of rigidity. Special scales were prepared because, in a previous experiment (2), predictions made on the basis of inventory scales of personality were not confirmed in a sufficiently clear-cut manner. It was felt that scale quality might have something to do with this, because in inventory scales the same score may be obtained from entirely different items, and it is not certain that the same score means the same degree of responsiveness or rigidity, for some items may be equivalent in intensity, some very different.

The *E* scale used here consists of the following items for males and females respectively:

Males

1. I tend to worry too long over humiliating experiences.
2. I have quite often had "stage fright" in my life.
3. I worry over possible misfortunes.
4. I am rather a nervous person.
5. I often feel disgruntled.
6. I am rather an irritable person.
7. I suffer from sleeplessness.
8. I suffer from attacks of diarrhea.

Females

1. I tend to worry too long over humiliating experiences.
2. I worry over possible misfortunes.
3. I am rather a nervous person.
4. I often feel disgruntled.
5. I have occasionally been troubled by stammer and stutter.
6. I am troubled by aches and pains.
7. I am subject to attacks of shaking and trembling.
8. I suffer from severe headaches.

The items on these scales are ranked from "mild" to "severe" degrees of anxiety, or emotional responsiveness. The scales are "cumulative," i.e., if a person checks positively an item he should also check positively all the lower ranked items. Since the scale reproducibility is .90 for males, and .92 for females, and *S*s are to some degree inconsistent, responses on one or two items may be "misplaced." The scales were constructed on students, and therefore are suitable only for student populations. Reapplications have been made both on students and on the population at large, however, and a rather remarkable reproducibility of between .84 and .88 was obtained.³

³ The detailed description of the scales construction will be published separately, with their new, revised versions in the September issue of *J. Pers.*

The preliminary R scale used in this experiment has twenty items (in two forms, one for males, one for females), and was constructed from *Pd* and *Pt* items of the MMPI. The rationale for combining these two types of items in one scale (2) is based on the hypothesis that psychopathic and compulsive personalities may be regarded as opposites in the following sense. Psychopathic personality is characterized by lack of generalized response tendencies, e.g., generalized adaptive responses, such as controls, which interfere with whatever situationally relevant, but possibly maladaptive, responses may be evoked. Compulsive personality is characterized by the presence of overlearned response tendencies, e.g., generalized maladaptive responses (defenses, or anankastic reactions), which interfere with situationally relevant, possibly adaptive responses. Most *Pd* and *Pt* items express this idea of interference. Therefore the hypothesis was formulated that those *Pd* and *Pt* items that express this idea of interference should scale in Guttman's sense, if *Pd* items are appropriately valued. The purpose of scaling was to rank them, if possible, according to the degree of interference each expressed. The items did actually scale, and therefore the scale was used to measure rigidity, that is, degree of interference of irrelevant responses with situationally relevant responses. Interference is interpreted as a loss of habit strength (2). Therefore, in order to avoid inequalities of habit strengths arising from this source, each pair of Ss was equated on rigidity scores. Since the variable measured by the R scale was controlled for all pairs, the items are not given here.³

Subjects

About 700 students in three general psychology classes were screened by the E and R scales. Thirty pairs were formed by taking two Ss with equally high R scores but different E scores (experimental pairs with low median E score of 1 and high of 6), and thirty pairs of Ss with equal R scores and equal E scores (control pairs). Such a selection of pairs was possible because the two scales proved uncorrelated for the sample chosen. Some very extreme cases with psychiatric backgrounds were omitted. All pairs were also equated as to sex, language spoken (English), nationality (Canadian), and course year (all three years of the General Arts Course were represented). The above characteristics define the two populations. Within these classifications Ss were paired at random.

Procedure

Three Ss participated in each session, giving three pairs each time. Two Ss had equal E scores, the third had a different one; all had equal R scores. This gave two experimental and one control pair. Toward the end of the experiment additional trios with equal E scores were obtained to get the required number of control pairs.

At the beginning of each session each S filled out a questionnaire on a series of topics previously obtained from students, stating both his opinion on each topic ("pro", undecided, or "con") and the strength of his feeling about the issue itself (very important, fairly important, or unimportant). Topics of very high emo-

tional value, such as religion, were omitted from the questionnaire. The questionnaires were then collected and, for each pair, a topic was selected on which the two members of the pair were of opposite opinions and on which both felt equally strongly (only "very important" topics were taken). The order of pairs was randomized to avoid any possible one-sided carry-over from experimental to control pairs, or vice versa. All questionnaires were then returned to their writers, and a pair instructed to discuss a given topic for 15 minutes or less, with the aim of reaching a consensus if possible. If no consensus was reached, the pair could disagree. The third S was instructed to act as chairman, keep track of time, and remain impartial. At the end of 15 minutes each S wrote openly on his questionnaire whether he agreed or still disagreed with his opponent. Qualified agreement was permitted as the third choice. The procedure was then repeated with the two remaining pairs.

Behavior Measurements

All sessions took place in a demonstration room separated from an observation room by a one-way vision screen. Voices could be transmitted between the two rooms as desired by the observers. There were two observers, one of whom was the experimenter who gave instructions to Ss.

1. Amount of an S's participation in the discussion was measured by adding the lengths of all his speeches, as recorded on the Interaction Chronograph. Reliability of this measure, obtained in previous experiments (2) by correlating two independent sets of observations, is .99.

2. Response latency, i.e., the time between the word "go" and the moment when each member of the pair made his first statement, was measured by a stop watch and also recorded on the I.C. Reliability of this measure is .99 (2).

3. Change of opinion was recorded by the observers, after the discussion, as "complete change," "qualified change," or "no change," in accordance with Ss' statements. This procedure avoided any possible mistakes of interpretation by observers, and was also necessary because of the absence of recording facilities.

MATHEMATICAL DERIVATION OF PREDICTIONS

Two classes of Ss, S_1 and S_2 are differentiated according to drive level: S_1 with high drive level (D_1), and S_2 with low drive level (D_2). Experimental pairs are formed by taking one S from class S_1 and one from class S_2 . Control pairs all have equal drive levels, and are taken at random from class S_1 or S_2 . Each S's response is classified, according to the opinion it supports, into one of two classes: his own or that of his opponent. The S defending his own opinion is designated by identical subscripts, e.g., S_{11} or S_{22} . If S makes statements supporting the opinion held by his opponent (his "alien" opinion), he is designated by different subscripts, e.g., S_{12} or S_{21} , the second subscript

standing for opinion class and the first for drive-level class. Thus H_{11} stands for the habit strength of S_1 's own opinion, H_{12} for that of his alien opinion; and similarly H_{22} stands for the habit strength of S_2 's own opinion and H_{21} for that of his alien opinion. It is assumed that these habit strengths vary during the experiment. It is further assumed that $H_{11} = H_{22}$, and $H_{12} = H_{21}$ at the beginning of the experiment, the first two being at their maximum, the second two at their minimum. It is also assumed that $H_1 = H_2 = \text{constant}$ (near unity) throughout the experiment, where $H_{11} + H_{12} = H_1$, and $H_{21} + H_{22} = H_2$. Therefore, when one of the two complementary habit strengths, e.g., H_{12} , increases, the other (H_{11}) is automatically extinguished (3). Equating S s on the importance of the issue under debate, and on the R score, was intended to contribute to equalization of the habit strengths H_1 and H_2 . Assumption of their constancy during experiments is supported by previous experimental evidence (2).

The reaction potentials E_{ij} and E_i ($i, j = 1, 2$) for S_i can be written as follows:

$$\text{and} \quad E_{ij} = D_i H_{ij},$$

$$\text{If} \quad E_i = \sum_j E_{ij}.$$

$$\text{then} \quad \sum_j H_{ij} = 1,$$

$$E_i = D_i.$$

The rate (p_i) of response of S_i per unit period can be written as a function of reaction potential:

$$p_i = f_i(E_i) = f_i(D_i), \quad [1]$$

and

$$p_i = \sum_j p_{ij} \quad [2]$$

where $p_{ij} = f_i(E_{ij})$. Equation 2 breaks down each S 's total response rate into two classes of responses according to opinions they support. The larger rate represents the dominant opinion.

If the rates of responding are measured as proportions of total speaking time of an S per unit-period, and if there are few total silences or simultaneous talking, $\sum p_i = 1$. With the definition of drive level adopted here and constant habit strengths H_i , both p_i 's should remain constant throughout the experiment.

This assumption is supported by the evidence of previous experiments (2).

Predictions of the amount of participation and latency are straightforward. The total amount of participation up to time T , measured in terms of the total length of time (t_i) spoken by S_i during that period is given by the cumulative rate equation

$$t_i = p_i T - a_i \quad (i, j = 1, 2). \quad [3]$$

where a_i is a constant and the rate p_i is the first derivative $p_i = dt_i/dT$. Substituting from [1] into [3] we get the prediction of the total amount of participation of either S :

$$t_i = f_i(D_i)T - a_i. \quad [4]$$

Since the length of the experiment is the same for both S s in a pair, the amount of participation t_i depends only on drive level, and should be larger for high- than for low-drive S s, within the limits of error.

If we think of the abscissa as the line representing the threshold of response, when $t_i = 0$, then

$$p_i T = a_i,$$

and the latency T_{Li} of S_i is

$$T_{Li} = a_i / p_i. \quad [5]$$

Substituting [1] into [5] we have

$$T_{Li} = a_i / f_i(D_i), \quad [6]$$

i.e., latency varies inversely with drive level, and therefore should be shorter for high- than for low-drive S s.

With allowance for error, predictions concerning latency and the amount of participation will be confirmed if, in the experimental pairs, either the mean $\bar{T}_1 > \bar{T}_2$, and $\bar{T}_{L1} < \bar{T}_{L2}$, or if there are more S_1 and S_2 with longer participation and shorter latency, as compared with control pairs, where no such differences should arise.

The derivation of predictions for change of opinion is as follows.⁴ If we think of change of

⁴ The writer is much indebted to Mr. V. Kahan, of the Department of Mathematics, University of Toronto, for assistance in the construction of this part of the model of interaction, without whose help the model would not be available at this time. He is also thankful to Dr. A. Robinson, of the Department of Mathematics, for helpful criticism of the mathematical part of the manuscript. However, the writer alone is responsible for any defects or mistakes in the derivation of predictions.

opinion as a learning process, then S_1 's responses constitute reinforcement for S_2 's response tendencies, and vice versa. The rate of S 's responding, and therefore reinforcement frequency, vary with time. The problem is therefore one of writing down the equations for variation with time of mutually dependent habit strengths and reinforcements. These equations are derived from the general learning function (3, 5).

The reaction potential as defined above ($E_{ij} = H_{ij}D_i$), is measured by p_{ij} . This can be written

$$E_{ij} = g_i(p_{ij}).$$

Substituting into the definition of reaction potential and solving for H_{ij} we have

$$H_{ij} = g_i(p_{ij})/D_i.$$

Dividing and multiplying the right-hand side by p_i , we have

$$H_{ij} = b_i h_i(r_{ij}),$$

where h_i is a monotonic increasing function, $r_{ij} = p_{ij}/p_i$, $0 < r_{ij} < 1$, $b_i = p_i/D_i = \text{constant}$, and $r_{i1} + r_{i2} = 1$.

The measure of the habit strength of S 's "considered" opinion, when "pros" and "cons" have been weighed, is defined as in a previous study (2) by

$$x_i = (r_{i1} - r_{i2}) \quad -1 < x_i < +1. \quad [7]$$

If $x_i < 0$, S_i mainly defends opinion 2; if $x_i = 0$, he is impartial; if $x_i > 0$, he mainly defends opinion 1. An increment (dx_i) of x_i as a result of discussion is

$$dx_i = (dr_{i1} - dr_{i2}).$$

Increment dr_{ij} is by definition (3, 5) made jointly proportional to the amount yet to be learned $(1 - r_{ij})$ and to the net increment (dn_{ij}) in reinforcement, i.e., the sum of reinforcement and extinction, received by S for each response tendency during the time interval (dT), if dT is sufficiently small so that within it all r_{ij} can be regarded as approximately constant:

$$dr_{ij} = c'_i(1 - r_{ij}) dn_{ij},$$

where c'_i is a constant, and dn_{ij} are the increments in reinforcement for the j th opinion. This equation will be recognized as the usual expression for the increment in habit strength (5).

The increment dx_i of x_i can therefore be written as

$$dx_i = c'_i[(1 - r_{i1}) dn_{i1} - (1 - r_{i2}) dn_{i2}]. \quad [8]$$

Reinforcement S_1 receives from his opponent is measured by the length of time during which the opponent speaks in support of one or the other opinion. The opponent's responses may affect either one of a silent S_1 's opinions, depending on which one of them was supported by the last overt or covert response. Therefore we have four possible effects of reinforcement on S_1 's opinions: reinforcement and "extinction" of opinion 1, and reinforcement and extinction of opinion 2. A similar statement can be made about S_2 when he becomes silent.

The increment of reinforcement or extinction received by S s in the interval of time dT can be estimated as follows: the interval dT is thought of as being split into eight portions: two S s speaking in support of two opinions, each response reinforcing or extinguishing a response tendency of the opponent. The reinforcement and extinction S s receive from their opponents, while they themselves are silent, are measured by the length of the fraction of the interval during which the opponent speaks in support of one or the other opinion. Since we cannot know what opinion S 's covert responses support while he is silent, the interval dT is divided in such a way that the last subdivision is into fractions according to the two classes of subvocal responses of the silent S . In virtue of $\sum_j r_{ij} = 1$, the reinforcement of one, and the extinction of the opposite response are equivalent. Both effects also are functions of the same responses of the opponent. This makes it unnecessary to know the magnitudes of covert responses of the silent S . With these considerations in mind, the net amount of reinforcement (dn_{i1}) of S_1 's opinion 1 is approximately given by

$$\begin{aligned} dn_{i1} &= r_{11}p_{21}dT - r_{11}p_{22}dT \\ &= r_{11}(p_{21} - p_{22})dT, \end{aligned} \quad [9a]$$

where $r_{11}p_{21}dT$ is the measure of the amount of reinforcement received by S_1 for opinion 1 during the fraction $p_{21}dT$ of the interval dT , when S_1 is silent, and S_2 supports opinion 1; $r_{11}p_{22}dT$ is the measure of the amount of extinction received by S_1 for opinion 1 during the fraction $p_{22}dT$, when S_1 is silent and S_2 supports opinion 2.

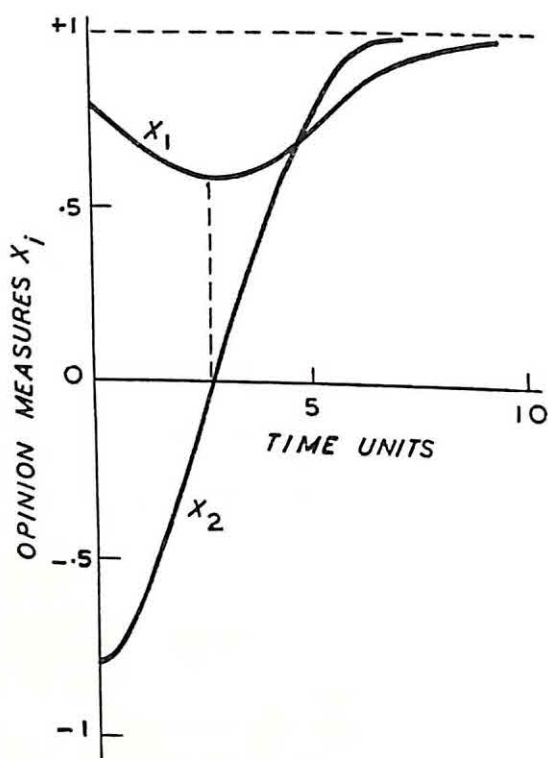


FIG. 1. NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS 11 AND 12 FOR EXPERIMENTAL PAIRS, SHOWING CHANGE OF OPINION OF LOW-EMOTIONAL S_2 WHO AGREE WITH HIGH-EMOTIONAL S_1

Similarly, the net amount of reinforcement (dn_{12}) of S_1 's opinion 2 is approximately given by

$$\begin{aligned} dn_{12} &= r_{12}p_{22} dT - r_{12}p_{21} dT \\ &= -r_{12}(p_{21} - p_{22}) dT, \end{aligned} \quad [9b]$$

where the interpretation of the terms is analogous to that in the preceding case. Equations 9 are simplified definitions of reinforcement measures, but more general functions can be written if required.

Substituting from [9a, b] into [8] we have for S_1

$$\begin{aligned} dx_1 &= c_1'[(1 - r_{11})r_{11}(p_{21} - p_{22}) dT \\ &\quad + (1 - r_{12})r_{12}(p_{21} - p_{22}) dT]. \end{aligned} \quad [10]$$

Using [7] and letting $c_1 = c_1'/2$ in [10], we obtain for S_1 after some rearrangement

$$dx_1 = c_1 p_2 (1 - x_1^2) \cdot x_2 \cdot dT, \quad [11]$$

and similarly for S_2

$$dx_2 = c_2 p_1 (1 - x_2^2) \cdot x_1 \cdot dT. \quad [12]$$

Constants c_1 and c_2 stand for effectiveness of S_1 's mutual reinforcement and are assumed to

be independent of time. To allow maximum communication efficiency, S_1 s were matched for age, sex, language (English), and level of education. Therefore the constants are assumed to have their maximum values of 1.

Equations [11] and [12] can be solved numerically,⁵ if p_1 , p_2 , x_1^0 , and x_2^0 (initial values), are known. The mean values p_1 and p_2 can be estimated from preadministered personality tests measuring reaction potential, i.e., drive level in our case. The estimates are $p_1 = .7$, and $p_2 = .3$ for the experimental pairs, and $p_1 = p_2 = .5$ for the controls. The mean initial values of x_1^0 and x_2^0 can be estimated from the preadministered opinion questionnaires. These estimates are based on actual rates of response in previous experiments for S_1 s with similar opinion scores. For both experimental and control S_1 s they are $x_1^0 = .8$, $x_2^0 = -.8$, with $|r_{ii}| = .9$. Numerical solution of Equations [11] and [12] for experimental pairs gives (a) an S shaped curve for x_2 as a function of T , which crosses the abscissa; (b) a trough-shaped curve for x_1 as a function of T , which does not cross the abscissa and whose minimum is at T where $x_2 = 0$. Both curves tend to the asymptote at $x_1 = x_2 = 1$. For the controls the two curves tend to the abscissa as their asymptote from the opposite directions. Thus in experimental pairs S_2 (low-drive S_1 s) should, as a result of the discussion, change their opinions to agree with S_1 , and in control pairs both S_1 s should yield some ground without actually switching their opinions. If the experiment lasts long enough for the process to run its course, and if we allow for error in the predictions, one should expect that in experimental pairs there would be more one-sided changes of opinion, with S_2 changing more often than S_1 . In control pairs, there should be more two-sided changes of opinion with few, if any, switches. This prediction is stated in frequency terms, so that the analysis can be made by nonparametric methods.

RESULTS AND DISCUSSION

Amount of Participation

Table 1 shows that the prediction of larger amount of participation for high-emotional S_1 s is confirmed at the .06 level of confidence.

⁵ Explicit criteria for predicting change of opinion can be derived from the model. Approximate solutions of the basic differential equations were also suggested by Dr. A. Robinson. Both will be published in a separate paper on the model.

TABLE 1

PARTICIPATION AND LATENCY DIFFERENCES
(+ AND -) BETWEEN HIGH- AND LOW-
EMOTIONAL Ss IN EXPERIMENTAL, AND
BETWEEN "FIRST" AND "SECOND" Ss
IN CONTROL PAIRS ($N = 60$)

Pairs	Participation Differences			Latency Differences		
	Posi- tive	Neg- ative	Totals	Posi- tive	Neg- ative	Totals
Experimental	25	5	30	7	23	30
Control	17	13	30	17	13	30
Totals	42	18	60	24	36	60
χ^2			3.57*			5.50*
p			<.06			<.02

* Corrected for continuity.

To aid in deciding whether this result should be taken as a confirmation of the hypothesis or not, it was also tested in a different way: the mean difference ($\bar{d}_e$) between the amount of participation of high- and low-emotional Ss in experimental pairs was $\bar{d}_e = +25.73$ periods of five seconds each (which was our measurement unit). This difference is significant at $p < .01$ level ($t = 4.1$). The mean difference between randomly numbered "first" and "second" Ss in the control pairs was $\bar{d}_c = -1.03$ periods, which is not significant ($t = .173$). The mean duration of the experiment with experimental pairs was 11 minutes, 54 seconds, with control pairs 11 minutes, 48 seconds; this difference is not significant. The results therefore support the hypothesis. In fact, the frequency test would also have come out significant, if a fifty-fifty break were taken as the null hypothesis, instead of the ratio obtained from the control pairs, which deviates slightly from an even division.

Latency of Response

Although latency was measured in seconds, the analysis was carried out by frequencies, since previous data (2) proved very difficult to transform in order to make the analysis by parametric methods possible. Table 1 shows that the hypothesis of shorter latency for high-emotional Ss is fully supported by the results.

Difference in latency of the first statement of opinion between experimental and control pairs was also analyzed, and gave a $\chi^2 = 4.03$ and $p < .05$ with 1 df , which means that latency of statement of opinion was also shorter

TABLE 2

FREQUENCIES OF ONE-SIDED AND TWO-SIDED CHANGES
OF OPINION IN EXPERIMENTAL AND
CONTROL PAIRS ($N = 60$)

Pairs	One-Sided Changes of Opinion	Two-Sided Changes of Opinion	Totals
Experimental	17	13	30
Control	5	25	30
Totals	22	38	60
p ($\chi^2 = 8.68^*$)			<.01

* Corrected for continuity.

for high- than for low-emotional Ss, as predicted. This is a more direct test of the prediction as derived, for the first response may be irrelevant to the issue.

Change of Opinion

The test of prediction of change in opinion was carried out in two steps: (a) first the hypothesis was tested that among the experimental pairs there would be more one-sided changes of opinion (complete or qualified) than among the control pairs; (b) then the hypothesis was tested that in experimental pairs that showed one-sided changes of opinion, more Ss who changed their opinions would have a lower, and fewer a higher, emotionality. Data in Table 2 support the first hypothesis.

To test the second hypothesis, the probability of no difference was calculated by the exact method for the 17 experimental pairs showing one-sided changes of opinion (see Table 2), among which there were four high-emotional and 13 low-emotional Ss who changed their opinions. The probability obtained is $p = .0245$ (1 tail). This supports the second hypothesis.

Discussion

The experiment reported here was designed in the first place to test the hypothesis derived from a model of free behavior of two interacting individuals. A few comments on the model seem necessary. The model as outlined in the mathematical section is a nonstochastic one. Tests of hypotheses show, however, that distribution functions were assumed to be associated with the parameters of the model. In this expanded form the model is stochastic (6), though somewhat different from the published models for one individual (3). Estes' model (3) can be derived from the present

model by keeping the reinforcement parameter of one S constant, i.e., by prescribing the role to be played by the "other" S . This kind of experiment was carried out by using "confederates" of the experimenter as S s (2).

The model presented here has the advantages of relative simplicity and of representing the social group as a whole consisting of inter-related parts. It can be generalized to any number of interacting persons and responses, but mathematical difficulties prevent this for the time being. All its parameters can be estimated before the experiment. Thus, curve fitting is not necessary if this model is used, for curves of performance can be traced beforehand, and empirical data compared with the theoretical (expected) values. The model can be used to predict not only the outcome of experiments, as has been done in this study, but also their course. In such cases, the functions obtained should be regarded as sort of a priori regression functions. The goodness of fit to data can be estimated by the usual statistical methods. The course of development of ascendance-submission has not been studied in the present experiment because of lack of technical facilities.

The model as outlined here is only one of the possible ones of the same type. By dividing the interval in different ways, other models could be generated. The advantage of dividing the interval dT in the way it was done here, is that no information is required about subvocal responses of an S while the other S speaks.

A secondary goal of the experiment was to test predictions based on measurements obtained by the new scales. Since predictions were confirmed in a clear-cut manner, it means that our scales are not worse than the inventory scales of personality measurement, and possibly better.

In conclusion it will be useful perhaps to point out the limitations of the present approach. First of all, ascendance-submission as operationally defined here is only one particular case of this relationship, for it can also occur under cooperation and other social conditions. Second, rather severe restrictive assumptions have been made. These have been tried out in previous experiments (2) and found necessary. It is not known, however, if they can always be experimentally approximated. Third, even the

model used here presents unusual mathematical difficulties: numerical rather than explicit solution of differential equations has been obtained, although approximate solutions are possible. However, if the prediction of change or no change in opinion is sufficient, explicit criteria for such a prediction can be derived from the model. If some of the restrictions were to be removed, purely mathematical research would have to be done before any further progress in the formalization of interpersonal interaction could be made.

SUMMARY

This study is an attempt experimentally to investigate the role of emotional responsiveness in the development of ascendant-submissive relationship in pairs of human individuals who discuss a topic on which they hold opposite opinions. The study was carried out from the point of view of stimulus-response theory of behavior. Specifically, it tested the hypothesis, derived from a model generalizing behavior theory to two individuals, that ascendant-submissive relationship between two verbally interacting subjects can develop as a function of their differential emotionality. Ascendant behavior of one subject over the other in a pair is defined in terms of larger amount of participation in the discussion, shorter latency of the first statement, and greater resistance to change of opinion. Emotional responsiveness, measured on a personality scale, is interpreted as drive level. The model used predicts that high-emotional subjects should tend to an ascendant, and low-emotional subjects to a submissive, position defined in terms of the three behavioral measures. Experiment was designed so as to form experimental pairs with subjects having different emotionality scores, and control pairs with equal scores. The three measures of behavior were analyzed by non-parametric methods. Results confirm the predictions: in experimental pairs, high-emotional subjects participated more in the discussion, had a shorter latency, and fewer of them changed their opinions. In control pairs, there were no significant differences in behavior. The model is stated in both verbal and mathematical form.

REFERENCES

1. BUCHWALD, A. M., & YAMAGUCHI, H. G. The effect of change in drive level on habit reversal. *J. exp. Psychol.*, 1955, **50**, 265-268.
2. CERVIN, V. Individual behavior in social situations: its relation to anxiety, neuroticism, and group solidarity. *J. exp. Psychol.*, 1956, **51**, 161-168.
3. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94-107.
4. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120-125.
5. HULL, C. L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951.
6. MARSCHAK, J. Statistical inference in economics. In Koopmans, T. C. (Ed.), *Statistical inference in dynamic economics*. New York: Wiley, 1950.
7. TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.

Received August 13, 1956. Prior publication.

A STUDY OF THE ADDITIVITY OF VARIABLES AFFECTING SOCIAL INTERACTION¹

BERNARD MAUSNER

*Graduate School of Public Health
University of Pittsburgh*

AND

BARBARA L. BLOCH

*Psychological Service
of Pittsburgh*

THE first stage in experimental study of social interaction was the demonstration that the phenomenon was observable in the laboratory (13). Succeeding investigations have related degree and occurrence of social interaction to specific variables. Consequently a great deal of information is now available on the parameters of social interaction (7). This work has prepared the ground for a third stage, one in which data on social interaction can be handled by means of a rigorous theoretical analysis, preferably quantitative. An example of such an approach is the recent attempt by French (6) to construct a miniature formal system embracing many aspects of social interaction.

One unstated assumption made by French, which is also the basis of other conceptual systems in this area (8, 12), is that the variables that have individually been demonstrated to affect social interaction are additive, i.e., that the effect of each variable on interaction is enhanced or diminished in a lawful way by the action of other variables. The factorial design is the most appropriate for testing the degree to which variables affecting social interaction show such additive properties. The present study was therefore planned with a

$2 \times 2 \times 2$ factorial design in which three variables, each with a high probability of affecting social interaction, were manipulated simultaneously. Two of the variables had previously been studied separately in a design similar to that of the present experiment: the degree of prior reinforcement of judgments received by each member of an interacting pair, and the demonstration by each partner to the other of success or failure in a previous judgment situation (presumed to affect each person's prestige for his partner). The third variable, presence or absence of previous experience in a cooperative task, had not previously been tested in precisely the situation employed in this study, but there was some indication in the literature (2, 6) that it should exert an effect on social interaction.

The experimental situation was that used in previous studies in this series (10, 11). Each S made a series of judgments without knowledge of his partner's responses. A series of judgments followed in which he was informed of the judgments purportedly made by his partner. The degree of social interaction is estimated from the amount and direction of shift in judgment between the first and second series of judgments.

The following hypotheses were tested:

1. Each variable affects the degree of social interaction in a predictable direction. That is, Ss who are positively reinforced in the first situation should converge less than those who are negatively reinforced; Ss with partners who are unsuccessful in the first situation should converge less than those whose partners are successful; Ss who had not participated in a cooperative activity prior to the experimental session should converge less than those who had participated in cooperative activity.
2. Behavior is most consistent among those groups of Ss for whom the influence of all variables is in the same direction. That is, Ss who are positively reinforced, have unsuccessful partners, and had not shared cooperative activity, should show the least convergence of

¹ This study is the fifth in a series entitled "Studies in Social Interaction." It was done under a grant to the senior author from the Dazian Foundation for Medical Research. Preliminary work under this grant was carried out at the University of Massachusetts, in facilities made available by the Department of Psychology. Mr. David Sands participated in the planning and execution of this stage of the work. The bulk of the project was done at the University of Pittsburgh by the junior author. The data formed the basis for an essay submitted to the University of Pittsburgh by the junior author in partial fulfillment of requirements for the master's degree. The writers wish to thank Dr. Harry W. Braun, Dr. A. W. Bendig, and Dr. Frederick I. Herzberg for their participation in discussions of the project. The grant was administered at this stage by Psychological Service of Pittsburgh, with which the senior author was affiliated at the time of the study. This assistance given by Psychological Service and its Director, Dr. Dora F. Capwell, is gratefully acknowledged.

all; those who are negatively reinforced for their first judgments, have successful partners, and had shared cooperative activity with their partners, should show maximum convergence.

3. Groups of Ss for whom two variables should enhance convergence and one should diminish show greater social influence than Ss for whom one variable should enhance convergence and two should diminish it. For these Ss, the degree of social influence on judgment should be intermediate between that in groups described in Hypothesis 2.

METHOD

Subjects. Eighty male undergraduate students at the University of Pittsburgh served as Ss. Each participant was arbitrarily placed in one of eight experimental groups. All Ss were naive with respect to the purposes of the experiment.

Stimuli. A film strip was prepared in which each of 35 frames consisted of a number of white dots on a dark background.² The strip was a negative print which had been made by photographing five times projections of each of seven slides containing varying numbers of dots. Each series of seven frames contained all seven slides arranged in a different order, so that no one slide followed any other more than once in the entire sequence. Evidence that the stimuli permitted wide variations in judgment is provided by the fact that estimates by various Ss on the frame with the largest number of dots ranged from 55 to 50,000.

Procedure. The Ss were scheduled to come to the laboratory in groups of two. In a preliminary briefing E indicated that the study was being carried out because of its similarity to a military situation.

One half of the eighty Ss was placed in groups for whom the first phase of the experiment consisted of the attempt by the Ss cooperatively to raise a ball to the top of a cardboard spiral track by manipulating the board on which the track was mounted. The task was introduced to the Ss by E as an approximation of a situation in which two men who do not know each other are required to perform a task which they have never seen before, a task which requires coordinated effort. Approximately ten minutes were allowed for the Ss to work on this task. A score, based on the height to which the ball was raised, was given as a reward for cooperative achievement. The remaining twenty pairs comprised the noncooperative group. For them this first phase was omitted.

For the second and third phases of the experiment, the two Ss, separated from each other by a partition, were seated facing a screen on which was projected the film strip described above. The E asked the Ss to think of each frame of the strip as a view of the sky with a number of jumping paratroopers in it. The job of the Ss was to estimate the number of parachutes in the sky.

In the second phase of the experiment, each of the 35 frames of the film strip was shown to the two Ss for

one-half second. After the exposure of each frame, each S wrote down his estimate and passed it back to E, who indicated aloud to S whether his judgment was wrong or right. A "right" judgment had previously been defined to the Ss as an estimate within 15 per cent of the actual number. Since E's report was audible to both Ss, each gained information as to his own and his partner's success. By giving false information (or, at least, information decided upon regardless of the S's actual estimate), both reinforcement vs. nonreinforcement and partner-success vs. nonsuccess variables were manipulated. The Ss in the reinforced group (R) were informed that they were "right" after 30 of the 35 frames; Ss in the nonreinforced group (NR) were told they were "right" on only 5 of the 35 trials.

If one of the Ss in a pair was in the reinforced group, his partner was, through the design of the second phase, made aware of this "success." That partner was, therefore, automatically in the "partner-success," or "high-partner-prestige" group (SP). A similar restriction obviously applies when one of the partners was in the nonreinforced group. For any individual S there were four possible combinations of these two variables: reinforced with a successful partner (R-SP), reinforced with a nonsuccessful partner (R-NSP), nonreinforced with a successful partner (NR-SP) and nonreinforced with a nonsuccessful partner (NR-NSP). Ten of the Ss given each of these four combinations of variables also experienced the cooperative task (C); the other ten did not (NC).

In the third phase, Ss were shown the same film strip in reverse, and were again asked to write down after each frame their estimates of the number of dots. They were told that this phase of the study was an attempt to simulate a situation where two platoon commanders can communicate with each other only by means of a messenger, and where the only information they can have as to the number of paratroopers is the estimate of the other person. Thus, E would not tell either S whether his estimate was right or wrong. Instead, after each trial, E took the slip of paper on which S had recorded his estimate, wrote on it a figure which ostensibly represented the estimate made by the other S, and returned the paper. Actually, the figure was a number which differed in a predetermined direction from the estimate made by S himself. For example, if in Phase II S tended to estimate in a range from 20 to 150, E returned to him in Phase III estimates that ranged from, say, 200 to 750. An attempt was made to keep the distance between S's estimates in Phase II and those purportedly made by his partner in Phase III proportionally equal for all Ss. The judgments of numerosity for the 35 frames shown in Phase III, as compared with his estimates in Phase II, constitute the data for each S.

Computation of convergence scores. The procedure used to assign to each S a score that would indicate the amount and direction of his shift from Phase II to Phase III is related to the technique previously used by the senior investigator (10, 11). This "D score" gives the frequency with which "numbers" used in the group-judgment situation differed from those used in individual judgments. These frequencies can be treated by means of parametric statistics. In the present case, to avoid negative numbers, a constant of 22 was added

² The stimulus was suggested by Dr. Solomon C. Goldberg and Dr. Ardie Lubin.

TABLE 1
CONVERGENCE SCORES FOR ALL SUBJECTS

+ + + C-NR-SP	- + + NC-NR-SP	+ - + C-R-SP	- - + NC-R-SP	+ + - C-NR-NSP	- + - NC-NR-NSP	+ - - C-R-NSP	- - - NC-R-NSP
70	66	49	53	72	1	32	34
71	37	63	2	39	68	31	43
77	80	18	42	36	43	60	55
81	76	47	26	56	63	30	34
72	19	41	47	62	43	47	40
76	67	40	28	24	21	25	37
80	74	71	53	38	39	7	28
56	55	77	20	49	41	46	27
76	44	41	53	24	13	40	33
76	46	76	82	39	8	41	26
Mean 73.5	56.4	52.3	40.6	43.9	34.0	35.9	35.7
SD 7.1	19.6	19.0	15.7	22.3	22.6	14.4	8.7

Note.—A *D* score of 22 involves minimal change in judgments from *alone* to *social* situations. *D* scores below 22 denote shift away from the partner, those above 22 shift toward the partner.

+ denotes a factor which was predicted as leading toward social influence, i.e., experience of cooperation (C), *S*'s own previous failure (NR) or partner's high level of success (SP).

— denotes a factor predicted as leading toward resistance to social influence, i.e., no cooperative experience (NC), *S*'s own previous success (R), or previous failure by the partner (NSP).

to all *D* scores. Thus a score of 22 represents no over-all shift in judgment; scores above 22 represent increasingly more frequent use by *S* of numbers similar to those purportedly used by his partner (i.e., *convergence*); scores below 22 represent *divergence* from the partner.³ Table 1 contains the *D* scores for each *S* and the mean *D* score and standard deviation for each of the eight experimental groups.

RESULTS

The main experimental variables. A $2 \times 2 \times 2$ analysis of variance was carried out. As the summary of this analysis in Table 2 indicates, the preliminary hypothesis that each of the three variables chosen would affect social interaction was confirmed. The findings relating to partner prestige and to *S*'s prior reinforcement corroborate the results of previous studies in which the effect of these variables on social interaction was demonstrated (3, 8, 10, 11).

The over-all effect of the ten minutes of cooperative activity engaged in by Group C in the present study was to increase the amount of social influence on judgments in the group-judgment situation. However, this finding holds only for the 30 *S*s for whom at least one

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE

Source of Variation	df	Mean Square	F
Between Success or Non-success of Partner	1	6,716.1	22.906****
Between Reinforcement Schedules	1	2,343.6	7.993***
Between Presence or Absence of Cooperation	1	1,891.5	6.451**
Interaction: Partner $\times$ Reinforcement	1	1,178.1	4.018*
Interaction: Partner $\times$ Prior cooperation	1	437.1	1.491
Interaction: Reinforcement $\times$ prior coop.	1	285.0	—
Interaction: Partner $\times$ Prior coop. $\times$ Rein.	1	23.2	—
Within Groups	72	293.2	—

**** Significant at the .001 level of confidence.

*** Significant at the .01 level of confidence.

** Significant at the .025 level of confidence.

* Significant at the .05 level of confidence.

other variable favored such influence (cf. Table 2, Groups C-NR-SP, C-R-SP, C-NR-NSP). Among the 10 *S*s in Group C-R-NSP for whom both of the other variables are opposed to convergence the effect of the cooperative experience was slight. One *S* showed considerable divergence from his partner's judgments; one converged greatly, the remainder have *D* scores within the same range as the *S*s who did not have the cooperative experience, and for whom the combination of prior success and low partner prestige led to main-

³ A detailed account of the procedure for calculating *D* scores has been deposited with the American Documentation Institute. Order Document No. 5092 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 3
TESTS OF HOMOGENEITY OF VARIANCE

Groups	χ^2	df	p
Total	18.242	7	.02-.01
C-NC	10.023	1	< .01
NR-R	0.060	1	.80-.90
SP-NSP	1.544	1	.20-.30

TABLE 4
F RATIOS OF THE VARIANCES OF EACH EXPERIMENTAL GROUP

	C-NR-SP	NC-NR-SP	C-R-SP	C-NR-NSP	NC-R-SP	NSP-NR-NC	C-R-NSP
NC-NR-SP	7.67**						
C-R-SP	7.15**	1.07					
C-NR-NSP	4.88*	1.56	1.46				
NC-R-SP	9.89**	1.28	1.38	2.02			
NC-NR-NSP	10.11**	1.31	1.41	2.06	1.02		
C-R-NSP	4.12*	1.86	1.74	1.18	2.40	2.45	
NC-R-NSP	1.51	5.07**	4.73*	3.23*	6.54**	6.69**	2.72

* Significant at better than the .05 level.

** Significant at better than the .01 level.

tenance of previous judgment norms. Thus for these Ss cooperative experience in itself did not overcome the effect of the other two variables in inhibiting social interaction.

Difference in variance among groups. Bartlett's test for homogeneity of variance was carried out. The resulting corrected χ^2 of 17.499 is significant at the .02 level. Despite this, the investigators accept the significance of difference among the means of the groups (Table 5) differing with respect to the three variables on the basis of statements by several statisticians (4, p. 166, 5, p. 129) that differences in means are almost never accounted for by lack of homogeneity of variance.

Two further approaches to the differences in variances among groups were made. Following a suggestion by Bendig,⁴ the Bartlett technique was applied to a comparison of sums of variances for the groups which differed with respect to each major variable and the χ^2 table entered with one degree of freedom. The χ^2 and p values are given in Table 3. According to this analysis, the lack of homogeneity of variance is partially due to differences between Ss who had had cooperative experience with the ball-and-spiral game and those who had not. The R-NSP group is the only one for whom cooperative experience did not lead to a considerable lessening in variability. As has

been noted, this experience did not, for most of the 10 Ss who had it, counteract the inhibition of social influence produced by the other two variables. The finding that a chance to participate with the partner in cooperative activity leads to a lessened tendency for Ss to vary in their response, and especially leads to less

TABLE 5
SUMMARY OF TESTS OF SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF EXPERIMENTAL GROUPS

Groups	t
SP and NSP	4.79***
R and NR	2.83***
C and NC	2.54**
NR-SP and R-SP	3.42***
NR-SP and NR-NSP	4.81***
NR-SP and R-NSP	5.39***
R-SP and NR-NSP	1.39
R-SP and R-NSP	1.97*
NR-NSP and R-NSP	.58

*** Significant at the .01 level of confidence.

** Significant at the .02 level of confidence.

* Significant at the .05 level of confidence.

likelihood of *divergence* is a phenomenon that deserves further study.

In the second approach F ratios were calculated comparing the variance of each experimental group with each of the others (Table 4). The results of this analysis are further discussed below.

Analysis of the components of significant interaction. One of the interactions in the analysis of variance was found to be significant, that between partners' prestige and reinforcement. Table 5 gives t tests showing the components of this interaction. The variables are related in a complex way. If the group of Ss who had successful partners is split on the basis of their level of prior reinforcement, a

⁴Personal communication.

significant difference is found. If the group with unsuccessful partners is similarly split, however, no significant difference emerges. The effect of both positive and negative reinforcement depends on the partner's prestige. Thus, for the groups as a whole, partner prestige and *S*'s prior reinforcement are additive, except in the situation in which *S* has a partner who has been unsuccessful. In that case, previous reinforcement does not significantly overcome the effect of the partner's failure to make "correct" judgments in the first situation. This result is consistent with common sense. There is little reason for *S* to converge toward a partner who has not demonstrated his competence in making judgments.

Comparison of the groups in which the influence of these two variables is in the same direction (NR-SP vs. R-NSP) shows a highly significant difference. In this case the two variables act cumulatively. The groups in which the variables are opposed (R-SP vs. NR-NSP) are not significantly different. Here the effects of the two variables presumably cancel each other out although actually the effect is to produce extremely high variance.

One of the most striking findings of the study was the fact that the two groups in which the influence of all variables lay in the same direction did not differ significantly in variance from each other but showed 11 significant *F* ratios out of the 12 possible relationships with the other six groups (cf. Table 4). In all cases, the variance was considerably lower for those groups in which all the variables had consistent direction. There were no significant *F* ratios among the variances of the groups in which the direction of the variables was opposed.

DISCUSSION

These findings have important implications for social interaction theory. The fact that the variance is significantly lower among the *S*s for whom the influence of all variables is in the same direction than among *S*s for whom the effects of the variables are in conflict implies that the concept of a resultant of forces—as used by French and others (6, 8, 12)—is limited in its usefulness. That is, such a resultant is meaningful only when the direction of the variables which produce the resultant is

consistent; where the direction of these variables is opposed, prediction of individual behavior from such an analysis is difficult.

Even the prediction of the behavior of groups of *S*s treated in the same manner is limited. The groups for whom prior reinforcement and partner success act in the same direction do differ significantly (cf. Table 5: NR-SP vs R-NSP). When all three variables are taken into account, however, the picture changes. It is true (cf. Table 1) that the three groups for which the influence of two of the variables is toward convergence (NC-NR-SP, C-R-SP, C-NR-NSP) all show higher average *D* scores than the groups for whom two of the variables inhibit convergence (NC-R-SP, NC-NR-NSP, C-R-NSP). However, the difference in over-all means between these two sets of groups is not significant ($t = 1.08, n = 58$), probably as a result of the very large variances. All of the "conflict" groups include both divergers and extreme convergers, as well as some *S*s whose judgments hardly shifted.

The dilemma of the *S*s in the NC-R-SP and NC-NR-NSP groups is especially great. These *S*s, who had had no experience with their partners outside the judgment task, were faced by a situation in which they had themselves been told that their judgments were wrong but had also heard their partners fail (NR-NSP) or had been told they were right and had heard their partners succeed (R-SP). For both groups, the inconsistent cues in the first experience were of little value to *S* in deciding how to react to the group-judgment situation. It is hardly surprising that some of these *S*s reacted by diverging, i.e., choosing a set of numbers differing both from their own and those purportedly used by their partners. Others hung on; still others yielded to their partner's influence.

A knowledge of *S*'s personality would probably be relevant in predicting which of these reactions he would show (1, 3, 9). In the present study the lack of variation among *S*s in the groups in which external determinants of interaction were consistent suggests that personality factors are least relevant where everything pushes *S* either to conform or to resist conformity. The great variation in response among *S*s in the "conflict" situations suggest that these create an ideal experimental

setting within which to examine further those characteristics of *S*'s personality which determine his susceptibility to social influence.

There is one further problem to solve before lawfulness can be demonstrated in the behavior of people in a social interaction experiment. A measure of the relative potency of the various factors affecting social interaction is necessary. There was no attempt in the present study to make such measures of the relative potency of the three factors. The high variance in the "conflict" groups suggests that for some *Ss* one factor was prepotent, for others another.

Thus two steps are required before the data from a social interaction experiment can be used as a basis for a mathematical theory in which behavior is derived from the resultants of forces. Firstly, personality measures are required for developing individual constants. Secondly, the major variables affecting social interaction must be manipulated systematically so that the relative effect of many levels of each variable can be estimated. When these steps are carried out the goals of theory construction will be more readily attainable.

SUMMARY AND CONCLUSIONS

The additivity of three variables found previously to be separately related to social interaction was investigated by manipulating them in a $2 \times 2 \times 2$ factorial design in a study of interacting pairs confronted with a task demanding perceptual judgment. The variables, treated dichotomously in the present study, were the individual's past history of success in making a given kind of judgment, his partner's corresponding history of success, and the presence or absence of previous experience on the part of the pair in working together on a cooperative task. Eighty male undergraduates, divided into eight experimental groups according to the factorial design, were tested in pairs. For each subject a score was computed showing the amount and direction of his shift in judgment when he was informed of the partner's alleged judgments. The findings may be summarized as follows:

1. Each of the three variables tested significantly affected the degree of social interaction in the predicted direction.
2. With the exception of the group in which

both of the other factors opposed social influence, one effect of prior cooperative experience was to decrease the variability of response among subjects.

3. Groups in which all three variables were predicted to affect interaction in the same direction (i.e., either in favor of or opposed to shift in judgment) showed the greatest consistency in response. Subjects in the group in which all variables favored social influence (cooperative experience, subject nonreinforced, high partner prestige) showed the greatest shift towards the partner; those in the group in which all variables opposed social influence showed the most consistent retention of previous judgments.

4. Those *Ss* for whom two variables pointed to convergence and one opposed it showed more social influence than the *Ss* for whom two variables opposed convergence. However, this difference was not significant. *F* tests showed significantly higher variance among these "conflict" groups than among the groups for whom the variables pointed in the same direction. Thus prediction of individual behavior in a social interaction situation is possible only when the factors affecting interaction point in the same direction. Where they are in conflict the responses of groups can be predicted to a certain extent, but the effect on individuals is not predictable. This finding was interpreted as a bar to the immediate testing of quantitative predictions based on theories of social interactions.

REFERENCES

1. ASCH, S. R. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men*. Pittsburgh: Carnegie, 1951.
2. BOVARD, E. W., JR. Group structure and perception. *J. abnorm. soc. Psychol.*, 1951, **46**, 398-405.
3. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191-198.
4. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
5. FISHER, R. A. *Statistical methods for research workers*. (6th Ed.) Edinburgh: Oliver & Boyd, 1936.
6. FRENCH, J. R. P., JR. A formal theory of social power. *Psychol. Rev.*, 1956, **63**, 181-194.
7. KELLEY, H. H. & THIBAUT, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 735-785.

8. KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267-285.
9. LINTON, HARRIET. Rorschach correlates of response to suggestion. *J. abnorm. soc. Psychol.*, 1954, **49**, 75-83.
10. MAUSNER, B. The effect of prior reinforcement on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 65-68.
11. MAUSNER, B. The effect of one partner's success or failure in a relevant task on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 557-560.
12. MAUSNER, B. Studies in social interaction: I. A conceptual scheme. *J. soc. Psychol.*, 1955, **41**, 259-270.
13. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, **27**, No. 187.

Received September 6, 1956. Prior publication.

EXTENT OF OPINION CHANGE AS A FUNCTION OF AMOUNT OF CHANGE ADVOCATED

CARL I. HOVLAND AND HENRY A. PRITZKER¹

Yale University

IN the literature on attitude and opinion change there are no clear-cut data available concerning the relationship between the amount of change in stand advocated by the communicator and the amount of change actually produced. Two diverse phenomena are characteristically obtained: (a) conformity to the recommendations of the communicator, with the result that the greater the change advocated the greater the expectation of change, (b) resistance to the communication, such that under certain conditions an adverse or "boomerang" effect is obtained. In the latter case, less change in the direction advocated by the communicator would be expected when greater change is advocated. In most prior studies the amount of change advocated is confounded with the position initially held by the recipient of the communication. Thus Sims (7) found that the more unfavorable the original opinion the greater the change produced when the communication was favorable to the issue. Similarly, Ewing (2) reports that the greater the difference between the subject's opinion and the position of the communication the greater the change brought about. An unpublished study by French and Gyr is cited by French (3) in which the amount of influence attempted was held constant. They found that the greater the discrepancy between the communicator and the communicatee the greater the change produced. Similarly, Goldberg (4) reports that the greater the discrepancy between the initial judgment of his subjects (in evaluating photographs) and the norm presented to them the greater the change. These studies deal essentially with the naturalistic correlation between communication-communicatee discrepancy and amount of

change produced rather than with the effects of experimental variation in the amount of change in opinion advocated. There is thus the possibility in some of these studies that there is a "ceiling effect" operative such that subjects with small discrepancies have less opportunity to change than those with larger discrepancies and also that subjects with more extreme position will on retest "regress" toward the mean. In the present study therefore the amount of change advocated was experimentally manipulated in such a way that small, medium, and large amounts of change could be advocated for subjects independently of their initial position on the issue.

METHOD

The over-all design of the study involved presenting to three groups of subjects three different communications, one advocating an opinion slightly different from the one held originally by the individual; one advocating a moderately different opinion, and a third a greatly different view. These three communications will be labeled "Slight," "Moderate," and "Marked" in terms of the amount of change in opinion advocated. Opinion questionnaires were administered and a month later a second questionnaire embodying the communication was given.

Because of the possibility of specific factors affecting the desired opinion change on any particular topic, twelve unrelated topics were chosen, each of which was embodied in a separate item in the questionnaire. The issues employed covered such diverse topics as the likelihood that electricity produced by atomic power would be available for use in American homes within five years and the desirability of making voting in presidential elections compulsory for all citizens over 21. A particular individual served in the "Slight" change condition on certain topics, in the "Moderate" change condition on others and in the "Marked" change condition for the remaining topics. Opinion on each of the twelve topics was obtained on a seven-point scale. The Ss were asked to check the one category which most nearly expressed their opinion toward the statement. After giving their opinion Ss were asked: "Of the following authorities which group's opinion would you respect most in reference to this question?" The names of three or four possible authority groups for the subject matter of the particular item were presented for choice. A sample question is given below:

1. a. All things considered, Washington was a greater president than Lincoln.

_____Agree strongly

¹ This study was carried out as part of the Yale Communication and Attitude Program under a grant from the Rockefeller Foundation whose generous support is gratefully acknowledged. A number of the questions used were standardized in a study by Dr. Bert King whose contribution to the study is gratefully acknowledged. The writers also wish to thank Drs. Harold Kelley and Fred Sheffield for helpful suggestions during the conduct of the study.

- _____ Agree moderately
- _____ Agree slightly
- _____ Undecided
- _____ Disagree slightly
- _____ Disagree moderately
- _____ Disagree strongly

b. Of the following authorities which group's opinion would you respect most in reference to this question?

- _____ Teachers
- _____ Historians
- _____ Parents

The answers to the second portion were used to determine which group would be used as the "communicator"; the group found to have the highest respect was used as the authority.

The second questionnaire, administered a month after the first, embodied the "communication." The procedure involved the use of check marks to indicate the position of the authority group. This method has been used by Arnett, Davidson, and Lewis (1) and by numerous later investigators. The second questionnaire contained the same opinion items as the first. However, beside each item the "opinion" of the chosen authority group was indicated. This opinion served as the communication and was varied according to the S's initial position on the issue.

What position was attributed to the authority was determined for each individual questionnaire on the basis of the responses obtained in the first questionnaire. One third of the Ss who gave a particular answer on the "before" questionnaire was given the "Marked" change communication, one third the "Moderate," and one third the "Slight." If, for example, the S's original opinion was "Agree strongly," a "Slight" communication consisted of the authority group's opinion being given as "Agree moderately"; a "Moderate" communication showed the authority position as "Agree slightly"; and a "Marked" communication indicated "Disagree slightly" for the authority. The patterns of communication for different initial positions are shown in Table 1.

The choice of communication S was to receive on each item depended upon his initial position with reference to that item. With this design the advocated opinions could be varied over the entire scale of posi-

tions from "Disagree strongly" to "Agree strongly," and the amount of advocated change could thus be made completely independent of the S's initial position.

In the instructions for the first questionnaire Ss were told that this was a study being conducted for the purpose of correlating students' opinions on various topics with their educational level. At the time of the second questionnaire Ss were told that the purpose of the second administration was to study the stability of opinion, and to make the study more interesting for them, opinions of various authorities as found by public opinion polls were shown. The fact that each of the second questionnaires was different necessitated giving each S a particular form. The Ss were told that this was being done to expedite the tabulating of results.

Fifty-one high school students who attended both sessions constitute the sample studied. The Ss were all in their senior year and represented the entire range of academic standing in the school.

RESULTS

Responses before and after the communication were scored on a seven-point scale from 7 for "Agree strongly" to 1 for "Disagree strongly." Difference scores were obtained by subtraction (After - Before). The direction of change was calculated from the standpoint of the position advocated so that plus changes indicate increased agreement with the communication and minus scores indicate increased disagreement ("boomerangs"). The mean scores for each type of communication are given in Table 2. It will be observed that communications that advocate a greater amount of change in fact produce a greater amount than communications that advocate a lesser amount. The amounts of change for the positive and for the negative communications are approximately the same. The tests of significance are based on the mean change scores for each individual for each category of change.

A separate analysis was made by individual

TABLE 1
POSITIONS ATTRIBUTED TO COMMUNICATOR IN RELATION TO INITIAL POSITION OF RECIPIENT FOR THE THREE COMMUNICATIONS ADVOCATING DIFFERENT DEGREES OF CHANGE

Subject's Original Opinion	Positions Attributed to Authority		
	Slight Change Communication	Moderate Change Communication	Marked Change Communication
Agree-strongly	Agree-moderately	Agree-slightly	Disagree-slightly
Agree-moderately	Agree-slightly	Disagree-slightly	Disagree-moderately
Agree-slightly	Disagree-slightly	Disagree-moderately	Disagree-strongly
Undecided	Agree-slightly or disagree-slightly	Agree-moderately	Agree-strongly
Disagree-slightly	Agree-slightly	Agree-moderately	Agree-strongly
Disagree-moderately	Disagree-slightly	Agree-slightly	Agree-moderately
Disagree-strongly	Disagree-moderately	Disagree-slightly	Agree-slightly

TABLE 2
MEAN CHANGE SCORE WITH THREE STRENGTHS OF
COMMUNICATION

	Degree of Change Advocated		
	Slight	Moderate	Marked
Positive communication	.89	1.33	1.57
Negative communication	.87	1.18	1.92
Combined	.88	1.25	1.75
	$p = < .03$		$p = < .05$
	$p = < .00001$		

TABLE 3
CHANGES IN OPINION WITH THREE STRENGTHS OF
COMMUNICATION SHOWN FOR EACH
INDIVIDUAL TOPIC

Topic	Degree of Change Advocated		
	Slight	Moderate	Marked
Washington or Lincoln greater President	.87	1.69	1.55
Married or single women better teachers	.93	1.25	1.85
Likelihood of cancer cure within 5 years	.87	1.00	1.58
Atomic power electricity for home by 1958	1.06	1.13	2.30
Adequacy of five hours sleep per night	1.54	1.72	1.42
Brains vs. beauty in choice of wife	1.30	1.64	1.69
Effectiveness of repetitive radio advertising	.75	.59	1.50
Sale of anti-histamines without prescription	.42	1.31	1.75
Compulsory voting by those over 21	.64	1.65	2.00
Number movie theatres still open by 1954	.81	1.50	2.00
Amount of power given President of U. S.	.78	.85	1.32
Russian superiority in number of A bombs	.64	.56	2.15
Median Change	.84	1.28	1.72

questions. Results for each topic are given in Table 3. The pattern shown in Table 2 holds for the individual items. The amount of change for the "Marked" change condition exceeded the "Slight" for 11 out of the 12 questions, the "Moderate" exceeded the "Slight" and the "Marked" exceeded the "Moderate" for 10 out of 12 questions.

The relationship between amount of change in opinion actually obtained and the amount of change advocated was found to be substantially the same for Ss holding the more ex-

TABLE 4
NUMBER OF "BOOMERANG" RESPONSES WITH THREE
STRENGTHS OF COMMUNICATION

	Degree of Change Advocated		
	Slight	Moderate	Marked
Number boomerang/ total responses	15/195	11/192	16/219
%	7.7	5.7	7.3

p_x^2 = not significant.

treme positions (strong or moderate agreement or disagreement) and for those holding more neutral positions (slight agreement or disagreement or undecided). Comparison of the *absolute* amount of change for these two groups is not possible without an additional control group, since regression effects might result in greater apparent change among those initially categorized at the extremes (cf. 6, pp. 329 ff.).

It was thought that there might be an increased tendency to give contra-communication or "boomerang" responses when the maximum change was advocated. However, no over-all differences for the three conditions were obtained (Table 4).

DISCUSSION

The results of the present experiment give clear-cut evidence that a greater change in opinion is produced by large than by small amounts of advocated change. Thus the results appear to be similar to those of Goldberg (4). They differ from his, however, in showing a smaller *relative* change the greater the amount of change advocated.² Thus when the communication supports one degree of change, the average amount of change produced is 88 per cent of that advocated; when two degrees are advocated, the proportion is 62 per cent; and when three are suggested, only 58 per cent of that advocated. But it is not possible to give a significance test for the differences between these percentages because of the nature of such ratios and because we cannot be sure we are dealing with a linear metric.

As implied at the outset we are not convinced that the same generalization is involved here as that given by French and others (3) to

² Robert Abelson has reanalyzed Goldberg's data and finds that the proportion of change in his study is not actually constant. Results for his intermediate group show a statistically significant departure from linearity.

the effect that the greater the deviation between the communicator and communicatee the greater the change produced. The generalization appears to us to involve two distinct factors: (a) the effect of amount of change advocated, and (b) the condition responsible for different individuals holding different positions initially. For some situations these two factors may both operate in the same direction, but for others they may cancel each other. Let us assume that the first factor operates as it did in the present experiment, i.e., the greater the change advocated the greater the change produced. The second factor can then work in the same or in an opposed direction. If the communication is kept constant and the correlation between discrepancy and change studied as in the French and Gyr studies, the second factor mentioned may mean that those with greatest discrepancies are initially at the more extreme positions. This circumstance may result in the extreme individuals being more resistant to change, producing an effect opposed to that produced by the first factor. On the other hand, if the communication is presented at one end of the scale, those already near this position have less possibility of change than those more remote; and this artifact ("ceiling effect") will increase the correlation between amount of discrepancy and amount of change. To avoid these problems the present study concentrated on an analysis of the first factor alone, and ensured that the communication was presented in such a way as to be independent of initial position. Furthermore, the position advocated was selected for each individual case in such a way that there was always equal opportunity for change, to avoid ceiling effects.

In considering the generality of the present results two considerations must be mentioned. First, in the present experiment the issues were not those involving deep-seated attitudes and beliefs. They would not be described as "ego-involving"; they were generally concerned with opinions that rest to a considerable extent on factual evidence. It would be expected that somewhat different results would be obtained with material in which the individual was deeply involved—there should be increased resistance when the individual's own position deviates markedly from that of the communicator.

A second factor to consider is the relationship between the communicator and the communicatee. In the present study the communicator was chosen so as to be someone respected by the recipient and hence an authoritative source of opinion. But with an untrustworthy communicator a quite different outcome would be expected. Similarly, with an unfamiliar or unidentified communicator a large deviation between the position advocated and that held by the recipient might serve to make him unacceptable. In many naturalistic situations one of the clues as to the acceptability of the communicator is the stand which he advocates. Thus a communicator advocating a political solution markedly more liberal than that held by the recipient may be labeled a "communist" with corresponding reduction in effectiveness. Under these circumstances advocacy of more extreme positions is likely to cause more resistance with occasional outright rejection of the communicator (cf. 5, pp. 41 ff.).

Research is needed to analyze the effect of these variables. By not identifying the communicator, except through his advocacy of one or another position, the interaction between the positions of the communicator and the recipient can be explored. An issue may be chosen which affects the individual deeply. The prediction is that under these circumstances an outcome quite different from that obtained in the present experiment will be produced. It is expected that under these conditions the greater the discrepancy between the position of the communicator and communicatee the greater the resistance to change, with actual adverse or boomerang effects predicted for the most extreme deviation conditions. An experiment is under way to test this prediction. The topic concerns prohibiting the sale of alcohol. The study is being made in states where prohibition is an actively debated issue, thus maximizing the amount of personal involvement.

SUMMARY

In previous experiments relating magnitude of attempted influence to degree of change produced, the initial position of the recipient and the amount of change advocated may have been confounded. In the present study three different degrees of change in opinion were advocated by communications presented in

such a way as to be independent of the recipient's own position. Opinions of Ss were secured on twelve different issues. A communication was then given a month later, consisting of transmission of information concerning the opinions held by respected authority figures on these issues. The Ss again expressed their own opinions. On one third of the issues the opinions of the communicator differed only slightly from that held by the S, on others somewhat more, and on still others quite markedly. The communications were hand tailored to the individual Ss by individually prepared checklists and questionnaires.

The communications advocating most extreme change had significantly more effect than those advocating moderate change ($p < .05$) and these in turn had significantly more effect than those advocating the smallest amount of change ($p < .05$). But the ratio of change produced to that advocated declined: the average amounts of change were 58 per cent, 62 per cent, and 88 per cent, respectively, of the amount of change advocated. The number of changes in a direction opposite to those advocated ("boomerangs") were substantially the same in each of the groups.

Theoretical considerations lead to the expectation that the positive correlation here obtained may be only generalizable to situations where a "positive" communicator (trustworthy, prestigious, well liked) is utilized. Different predictions would be made for "negative" and neutral communicators.

REFERENCES

1. ARNETT, C. E., DAVIDSON, H. H., & LEWIS, H. H. Prestige as a factor in attitude change. *Sociol. & soc. Res.*, 1931, **16**, 49-55.
2. EWING, T. N. A study of certain factors involved in changes of opinion. *J. soc. Psychol.*, 1942, **16**, 63-88.
3. FRENCH, J. R. P., JR. A formal theory of social power. *Psychol. Rev.*, 1956, **63**, 181-194.
4. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325-329.
5. HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.
6. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. Princeton: Princeton Univer. Press, 1949.
7. SIMS, V. M. Factors influencing attitudes toward T.V.A. *J. abnorm. soc. Psychol.*, 1938, **33**, 34-56.

Received November 19, 1956. Prior Publication.

CRITIQUE AND NOTES

THE TAYLOR SCALE AND REACTION TIME¹

LEON J. KAMIN AND JAMES W. CLARK

Queen's University²

THE study reported here explores the relations between Taylor's scale of manifest anxiety (7), simple reaction time (SRT), and reaction time motivated by avoidance of shock (ART). The existing evidence in this field is confused. Thus, Wenar (8) found a high manifest anxiety score (A score) to be associated with quick SRT; most of his Ss, however, were reacting to an electric shock stimulus. The studies by Farber and Spence (1, 2), one of which (2) required all Ss to react to shock, found no effect of A score on SRT. The threat of shock for reacting slowly failed to affect high and low A scorers differentially (1). The threat was bogus, however, as no shocks were in fact given.

The interpretation of the cited experiments is clouded by two alternate interpretations of A score espoused by the same writers (1, 2, 6, 8). The A score may be regarded as indexing a drive level (*D*) which (*a*) is characteristic of *S* in all situations or which (*b*) is characteristic of *S* only in situations which elicit anxiety. The latter interpretation suggested the experimental design employed in this study.

The same Ss were tested both for SRT and for ART, with the SRT trials kept completely free of threat. The view that A score predicts *D* in all situations suggests that A score should be associated both with SRT and ART. The alternate view suggests that A score should be associated with ART but not with SRT. This notion also implies that A score should relate most sensitively to a ratio or difference measure which contrasts S's ART and SRT. This type of measure serves to cancel out all variables which affect experimental performance other than the effect on *S* of changing from threat-free to avoidance-of-shock conditions. This effect on *S*, of course, is in essence what the second interpretation of the A score believes the A score to predict.

METHOD

Subjects and Apparatus

The Ss were 67 normal adult males in basic training with the Canadian Air Force. They ranged in age from 17 to 28 and had completed from 8 to 12 years of school. The apparatus consisted of a common buzzer, a

telegraph key, and a Standard Electric chronoscope. The *E* activated the buzzer and chronoscope simultaneously; *S* stopped both by pressing the key.

Prior to the ART trials, finger-clip electrodes wired to an inductorium were attached to two fingers of *S*'s nonpreferred hand. The inductorium, powered by a 6-v. dry cell, delivered shock whenever *S* failed to press the key within 180 msec. after onset of the buzzer. This interval was controlled by a Hunter timer. The *S* sat before a screen which masked all apparatus but the key and electrodes.

Procedure

SRT trials. When *S* was seated, *E* indicated the key and instructed *S* to press it as quickly as he could whenever he heard the buzzer. There were 3 unrecorded practice trials, followed by 18 SRT trials. The intertrial interval varied randomly between 5 and 12 seconds. There was a verbal ready signal by *E* 2 seconds before each buzz.

Shock adjustment. The finger electrodes were then brought into view for the first time and attached. The *S* was instructed to select a level of shock which was "really uncomfortable without being really painful."

ART trials. The *S* was then instructed to continue responding to the buzzer as quickly as he could. This time, he was told, he would be shocked whenever he responded "too slowly." The *S* was in fact shocked whenever his latency exceeded 180 milliseconds. The shock, when delivered, was terminated by *S*'s pressing the key. There were 18 ART trials with the same intertrial intervals as were employed during SRT trials.

Measures

The reaction time for each trial, recorded to the nearest .01 second, was transformed to its reciprocal, thus providing a speed-of-reaction measure. The ratio and difference measures were computed for each *S* by contrasting his mean ART and SRT speeds of reaction. The Taylor Scale, without filler items, had been administered to each *S* to provide an A score.

RESULTS

The product-moment correlations between mean SRT and ART speed-of-reaction scores, ratio and difference scores, and A score are given in Table 1.³ These data are surprising. There are significant relationships between A score and both SRT and ART, but these relationships are in a direction

³ The data were also analyzed with rank-order correlations employing the median rather than mean raw SRT and ART values. The coefficients obtained in this way very closely approximated those reported and in no way change the interpretation.

¹ Research done under the auspices of D.R.B. Grant 9435-55, PCC # D77-94-35-46.

² The research was carried out at McGill University.

TABLE 1
PRODUCT-MOMENT CORRELATIONS AMONG
VARIABLES
N = 67

Correlation	Coefficient	<i>p</i>
A-score $\times$ SRT	-.44	< .01
A-score $\times$ ART	-.24	< .05
A-score $\times$ (ART/SRT)	.32	< .01
A-score $\times$ (ART-SRT)	.34	< .01
ART $\times$ SRT	.68	< .001

opposite to that expected if A score is regarded as a measure of *D*! That is, a high A score is associated with low speed of reaction under both experimental conditions.

The relations between A score and the ratio and difference measures are more in keeping with theoretical expectations. These measures are significantly associated with A score. The higher the A score, the greater the increase of speed of reaction from SRT to ART conditions. The same tendency is shown by the fact that the correlation coefficients between speed of reaction and A score obtained under SRT and ART conditions differ significantly from each other ($p < .05$).

The ART speed of reaction was greater than the SRT speed of reaction. This was true of 59 of the 67 Ss ($p < .001$). While the difference between SRT and ART is confounded with order, there was little evidence of improvement *within* either the SRT or ART trials.

The high correlation between SRT and ART emphasizes the limit placed by the nature of the task upon any correlation between an independent measure of anxiety and ART performance. That is, ART performance is largely determined by factors quite independent of stress or anxiety.

DISCUSSION

Perhaps the simplest interpretation of the data is that A score is confounded with some other variable which is inversely related to speed of reaction. The hypothesized other variable, however, is unknown. There is much evidence that, at least among military populations, A score is significantly associated with various "intellectual" variables (3, 4, 5), but there is no evidence that such variables are related to SRT or ART.

Whatever the confounding variable may be, there is evidence that A score—among other things—does predict the effect upon *S* of the introduction of stress into an experimental task. This is suggested by the significant relations between the ratio and difference scores and A score. This effect could not have been noted had Ss been tested under ART conditions only; such a test would have sug-

gested merely that A score was inversely related to ART speed of reaction.

The inverse relations between A score and speed of reaction might be interpreted by assuming that anxiety has a disorganizing rather than a facilitating effect even upon such simple performances as reaction time. This view, however, fails to suggest why anxiety should have a significantly *less* disorganizing effect upon ART than upon SRT!

The various discrepancies between these data and those of Farber and Spence (1, 2) and Wenar (8) cannot now be satisfactorily interpreted. Perhaps the differences in procedure, the use of military rather than college Ss, or sheer sampling error are responsible. The data as a whole suggest more than anything else the need for extreme caution before interpreting correlations between A score and experimental performances as relevant to theoretical propositions concerning the nature either of *D* or of anxiety.

SUMMARY

The same 67 Ss were tested for simple reaction time (SRT), reaction time motivated by avoidance of shock (ART), and manifest anxiety score (A score). The higher the A score, the slower was both SRT and ART. The higher the A score, however, the greater was the increase in speed of reaction from SRT to ART conditions.

REFERENCES

1. FARBER, I. E., & SPENCE, K. W. Main and interactive effects of several variables on reaction time. Iowa City: State Univer. of Iowa, Mar. 1, 1955. (ONR Contract N9 onr-93802, Tech. Rep. No. 3.)
2. FARBER, I. E., & SPENCE, K. W. Relation of stimulus intensity, practice, anxiety, and sex to simple RT and temporal generalization. Iowa City: State Univer. of Iowa, June 27, 1955. (ONR Contract N9 onr-93802, Tech. Rep. No. 4.)
3. GRICE, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, 50, 71-74.
4. KAMIN, L. J. Relations between discrimination, apparatus stress, and the Taylor scale. *J. abnorm. soc. Psychol.*, 1955, 51, 595-599.
5. KERRICK, J. S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1955, 50, 75-77.
6. SPENCE, K. W., FARBER, I. E., & TAYLOR, ELAINE. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, 48, 404-408.
7. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
8. WENAR, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, 49, 335-340.

Received March 8, 1956.

GENERALITY OF SPEED AND CONFIDENCE OF JUDGMENT¹

LAVERNE C. JOHNSON²

Stanford University

A SURVEY of the literature indicates that the consistency of speed and confidence of judgment from task to task is still an unanswered question (1, 2, 3, 4). The generality of speed and confidence is of first importance because if these variables are of limited generality, reference to them must be rigorously qualified as to task and situation. If a relatively high degree of generality can be demonstrated, then speed and confidence become potentially useful dimensions in the study of judgment and choice.

METHOD

Subjects (Ss) were 71 male Stanford students enrolled in an elementary psychology course. The mean age of the sample was 20.8 years.

In planning the experiment, particular attention was given to problems of examiner influence, involvement and set of S, selection of the tasks, and variability from one session to another. Most previous research on speed and confidence of decision has involved objective psychophysical tasks that usually called for two-category judgments. To broaden the area of study, subjective, aesthetic, and objective tasks were included. Judgments were required not only between two stimuli but also among several stimuli.

To measure the degree of confidence a continuous scale was constructed with five defined points, including the two extremes of A, "100 per cent sure", and E, "a mere guess." After each judgment, S referred to the scale and gave the letter A, B, C, D, or E depending on his confidence in the adequacy of his response. A quantified score was obtained by giving the "A" response a score of five, the "B" four, "C" three, "D" two, and "E" a score of one.

Decision Tasks³

Picture preference. This task consisted of 30 black and white art reproductions which the S was to arrange

in order of preference from the most liked to the least liked using any scheme he wished.

Brightness matching. The S was told that "we wanted to find out something about the brightness of a tone," and he was to decide whether each of a series of lights appeared to be more brilliant, equal to, or less brilliant than a constant tone.

Recognition of figures. At the end of the session, S was required to decide which of 36 designs he had seen at the beginning of the hour.

Line judgment. The S was required to judge whether one line was shorter or longer than another.

Judgment of position of hand. The S was blindfolded and then asked to decide by running his hand along a meter stick whether he had passed certain points.

Weight judgment. The S was blindfolded and then asked to decide whether a series of weights were heavier, equal to, or lighter than a standard weight.

Angle judgment. This task was a modification of that used by Wolff (5) in his investigation of certainty.

Object sorting. Seventy-three objects were to be put in as many groups as S wished so long as all the objects in each group belonged together for a particular reason.

RESULTS AND CONCLUSIONS

In Table 1 are the means, medians, and standard deviations of the speed and confidence of judgment scores. The speed of decision scores were markedly skewed and a log transformation was made to normalize the distribution. All correlations of time scores were computed using log scores.

Table 2 gives the reliability of the various speed and confidence measures. Because of the nature of the picture-preference and object-sorting tasks, it was not possible to ascertain the reliability of these measures. Unless otherwise indicated, reliability was evaluated by the odd-even split-half correlational method corrected by Brown-Spearman formula for length.

Generality of speed. In Table 3 are the intercorrelations of the seven speed-of-judgment scores and the average correlation of each task with the other six. The r to z transformation method for averaging correlations was used. The mean of the 21 correlations is .48. All the correlations except one, .26, are significant beyond the .01 level, the requisite level accepted for statistical significance and indication of generality throughout this study. The correlations of .26 between picture preference

found in a copy of the dissertation obtainable from University Microfilm, 313 N. First Street, Ann Arbor, Michigan. Publication Number is 9501. The dissertation is on file at Stanford University Library under the title, "An Investigation of Speed and Confidence of Judgment as Psychological Variables."

¹ This report is part of a dissertation submitted to the Department of Psychology and the Committee on Graduate Study of Stanford University. The writer wishes to express his appreciation to Dr. Clarence L. Winder, Dr. Quinn McNemar, and Dr. Kenneth B. Little for their generous help in this research. This study was carried out under Project NR 150-087 supported by contract Nonr 225(01) between Stanford University and the Office of Naval Research. Work on the contract is under the general direction of Dr. C. L. Winder. Permission is granted to the United States Government for reproduction, translation, publication, use, and disposal of this article in whole or in part.

² Now at Washington University School of Medicine, Department of Psychiatry and Neurology, St. Louis, Missouri.

³ Instructions and material for measures may be

TABLE 1
QUANTITATIVE DESCRIPTION OF MEASURES

Task	Speed of Judgment*			Confidence of Judgment		
	Median	Mean	SD	Median	Mean	SD
Picture preference	393.3	455.4	173.8	—	—	—
Object sorting	379.5	369.9	134.8	—	—	—
Angle judgment	309.5	354.6	152.7	143.7	144.0	17.3
Brightness matching	119.0	138.4	78.4	147.9	148.5	16.7
Line judgment	152.8	173.9	100.2	143.0	137.8	23.0
Weight judgment	324.5	370.2	185.1	139.9	138.8	24.0
Figure recognition	123.7	143.0	76.6	138.8	138.5	19.0
Position of hand	—	—	—	136.0	141.5	20.2

* All time in seconds.

TABLE 2
RELIABILITY OF MEASURES OF SPEED AND CONFIDENCE

Task	Speed		Confidence	
	Reliability coefficient	N	Reliability coefficient	N
Angle judgment	.96*	24	.93	40
Brightness matching	.96	50	.96	40
Line judgment	.87**	32	.82**	32
Weight judgment	.96	50	.96	40
Figure recognition	.96	50	.93	40
Position of hand***	—	—	.97	42

* Evaluated by split half.

** Test-retest between sessions.

*** Speed-of-judgment score was not recorded.

TABLE 3
INTERCORRELATIONS AMONG MEASURES OF SPEED OF JUDGMENT

	Picture Preference	Angle Judgment	Brightness Matching	Line Judgment	Object Sorting	Weight Judgment	Figure Recognition
Pictures							
Angles		.62					
Brightness			.35				
Lines			.50	.37			
Objects				.52	.54		
Weights				.47	.37		
Figures				.72	.30	.31	
Average	.42	.46	.51	.53	.46	.47	.49

and figure recognition is significant at the .02 level, almost reaching acceptable significance. These results indicate that speed of judgment is significantly associated from one task to another under the conditions of this study.

Generality of confidence. Intercorrelations of the six confidence-of-judgment scores are presented in Table 4 with the average correlation of each measure with the other five. The mean of the 15 correlations is .67. All the correlations are well beyond

TABLE 4
INTERCORRELATIONS AMONG MEASURES OF CONFIDENCE OF JUDGMENT

	Angle Judgment	Brightness Matching	Line Judgment	Position of Hand	Weight Judgment	Figure Recognition
Angles		.73	.57	.58	.47	.53
Brightness			.71	.71	.66	.58
Lines				.67	.70	.72
Position of hand					.78	.71
Weights						.77
Figures						
Average	.58	.68	.68	.69	.69	.67

the .01 level accepted as indicative of generality. The obtained degree of relationship indicates that confidence of judgment is less task-determined than speed of judgment.

Whether speed and confidence were related to accuracy was not directly investigated. However, to see whether the generality of speed and of confidence could be attributed to a general accuracy factor, the accuracy scores on the tasks were intercorrelated where such a score was feasible. The range of correlations was from $-.13$ to $+.33$ with an average correlation of .10. There was no evidence of a general accuracy factor, and the possibility that accuracy was an important contributing factor to the generality of speed and confidence was not supported.⁴

A negative correlation of .38 was found between total speed and total confidence.

REFERENCES

- BRIDGES, J. W. An experimental study of decision types and their mental correlates. *Psychol. Monogr.*, 1914, 17, No. 1 (Whole No. 72).

⁴ As part of the dissertation, measures of manifest anxiety, need achievement, ego control, and authoritarianism were related to speed and confidence of judgment. No significant relationships were found.

2. FILTER, R. O. An experimental study of character traits. *J. appl. Psychol.*, 1921, 5, 297-317.
3. JOHNSON, D. M. Confidence and speed in the two category judgment. *Arch. Psychol.*, 1939, 34, 1-52.
4. TROW, W. C. The psychology of confidence. *Arch. Psychol.*, 1923, 10, 1-47.
5. WOLFF, W. M. Certainty: Generality and relation to manifest anxiety. *J. abnorm. soc. Psychol.* 1955, 50, 59-64.

Received March 12, 1956.

THE EFFECTS OF INTERCULTURAL EXPERIENCE— A FOLLOW-UP INVESTIGATION¹

HOWARD PERSIFOR SMITH

Bennington College

As Smith has pointed out in a careful analysis of research in the intercultural field (7), one of the unexplored problems concerns the long-run effects of intercultural experience. Of the published follow-up studies, the Watson & Lippitt investigation of German leader teams six to eight months after their return to Germany (9) is the only longitudinal research, Riegel's study of Belgians who had been to the United States (3) being based on temporal cross sections of former grantees. There has been no long-range study of Americans who have had intercultural experience. As a first step in this direction the present paper reports a follow-up investigation of the attitudes and internationally-oriented behavior of 151 young Americans who had previously been studied (6) immediately before and after spending the summer of 1950 in Europe.

METHOD

Early in May, 1955, precisely five years after they had first been tested, a mail questionnaire was sent to 215 subjects (Ss) in four experimental groups, two posttest-only control groups, and a "stay-at-home" questionnaire. The composition of the various groups is described in the original article (6, p. 469). After two follow-up communications, 156 Ss (72.6 per cent of the mailing group) returned completed questionnaires. Table 1 indicates the number of Ss in the various groups and the rates of response.²

¹ This study was made possible by a grant from the Edward Yarnall Hartshorne Memorial Fund of The Experiment in International Living. I am indebted to Miss Kay Crawford for her assistance in carrying out the study.

² Unpublished findings by the writer (5) lead to the expectation that the subjects who did not respond would be less worldminded and more ethnocentric than the respondents. The mean 1950 posttest scores on the Worldmindedness and Ethnocentrism scales of the 1955 respondents and nonrespondents differ in the expected direction. In the case of the W scale the difference is not significant by a one-tailed test ($.05 < p < .10$, $t = 1.29$),

TABLE 1
CHARACTERISTICS OF SAMPLES

Sample	Number returning 1950 posttest	Number receiving 1955 follow-up	Number returning 1955 follow-up	Percentage returning 1955 follow-up
Experiment*	105	102	75	73.5
NSA†	36	34	24	70.6
Tourists	48	45	35	77.8
QIVS‡	26	24	17	70.8
Stay-at-home control	11	10	5	50
All groups	226	215	156	72.6

* The Experiment in International Living.

† National Student Association.

‡ Quaker International Voluntary Service.

The mail questionnaire contained (a) the same Worldmindedness-scale (4), Democracy-scale (2), and E-, F-, and PEC-scale items (1) as appeared in the 1950 pretest and posttest questionnaires, (b) the same questions about the Ss' internationally-oriented behavior, and (c) several new questions about the Ss' 1950 experience and about their activities since 1950.

RESULTS AND DISCUSSION

Association with Europeans and Changes in Behavior

The original finding that Americans who had established close personal ties with Europeans were more active in certain internationally-oriented behavior was based on significant differences in the fall of 1950 (but not in the spring) between Ss who as members of The Experiment in International Living (Experimenters) lived in European homes and Ss who travelled under the auspices of the National Student Association, the Quaker International Voluntary Service, or as tourists, in amount of European correspondence and number of persons sending gifts to Europe. Four and a half years later Experimenters continue to be more active than the other Ss in their international be-

but the difference is significant in the case of the E scale ($p < .05$, $t = 1.82$).

havior according to these two measures: with respect to number who correspond with Europeans, the greater proportion of Experimenters who engage in this activity is still significant ($p < .02$, $t = 2.54$), but the proportional difference in number of Ss sending gifts falls below significance ($.10 < p < .15$, $t = 1.52$).³

Three other measures of international activity in the 1950 study were membership in organizations with primarily international interests, choice of vocation in some phase of international affairs, and college major in international studies. In 1950 neither significant changes in these activities over the summer nor significant differences among the experimental and control groups were found. The 1955 results corroborate the 1950 findings. Not only has there been no relative change among groups in these activities, but also the absolute number of persons reporting any of the activities remains very low; neither in the spring of 1950 nor in either of the two follow-up periods did as many as 5 per cent of the Ss in any group list activities primarily concerned with some phase of international affairs.

Changes in Social Attitudes

The short-run investigation found no significant changes in general social attitudes during the four-to six-month testing interval. Four and a half years later several significant changes in attitude appear. The data are given in Table 2. In the case of the two attitudes relatively more salient to an intercultural experience, there was a significant decrease both in worldmindedness and in ethnocentrism for the Ss who had been to Europe ($p < .001$ for the W scale, $t = 4.30$; $p < .001$ for the E scale, $t = 3.91$).⁴ Evidence suggesting that neither of these changes is a delayed outcome of the intercultural experience itself comes from the stay-at-home control-group findings. Although the loss of cases from this sample has seriously impaired its usefulness as a control group, it is worth noting that the control group changes on the W and E scales are in the same direction as the European group changes; the E-scale change just reaches significance for four degrees of freedom ($p < .05$, $t = 2.78$), but the W-scale change is not significant ($.10 < p < .20$, $t = 1.78$).

³ Since 56% of the Experimenters had been "outside North America" after 1950 compared to only 43% of the other Ss, and since the data revealed that Ss irrespective of group affiliation who had been abroad after 1950 were more likely to correspond and to send gifts, only those who had remained in North America were compared.

⁴ Although, as Footnote 2 implies, the 1955 W scale mean would have been smaller and the E scale mean larger with a 100% return, there is no reason to believe that the amount of change in attitude on the two scales is affected.

TABLE 2
ATTITUDE SCALE MEANS

Attitude Scale	Date Tested	European-Travel Groups*		Stay-at-Home Control Group*		Intensive Contact Group	Little Contact Group
		R	NR	R	NR		
N		151	183	5	11	12	18
Worldmindedness	5/50		126.73		126.27	137.17	129.00
	10/50	130.79	128.35	127.80	128.27	134.92	130.44
	5/55	123.66		110.60		132.50	124.89
Ethnocentrism	5/50		23.94		21.91	21.00	22.33
	10/50	21.92	23.16	20.60	21.18	19.83	22.61
	5/55	20.37		15.00		18.75	20.50
F	5/50		43.40		38.64		
	10/50	42.08	42.47	32.40	37.36		
	5/55	38.15		30.80			
Political-Economic Conservatism	5/50		37.75		40.54		
	10/50	37.64	38.00	36.20	37.64		
	5/55	40.11		42.40			
Democracy	5/50		33.11		34.73		
	10/50	32.68	32.95	32.80	33.82		
	5/55	32.39		35.40			

* R indicates 1955 respondents; NR indicates 1955 non-respondents.

No change appeared in the Democracy-scale means, but a highly significant change in attitude was found on both the F and Political-Economic Conservatism scales. The Ss were less authoritarian in 1955 than in the fall of 1950 ($p < .00001$, $t = 6.86$) and more conservative ($p < .00001$, $t = 5.09$). The F-scale change for the four European groups is not paralleled by the stay-at-home control group, but the greater conservatism of the control Ss on the PEC scale is in the same direction as the experimental group change ($.10 < p < .15$, $t = 2.06$).

The data are equivocal as to the source of these long-run changes in attitude. However, it does appear that events since 1950—either changes in cultural ideology or psychological changes associated with the Ss' change in age status from late adolescence to young adulthood—have had a far greater impact on their social attitudes than the intercultural experience they shared.⁵

⁵ In regard to the 17 Ss singled out for special study in the original investigation because they showed the greatest combined change in score on the W and E scales in the direction of becoming more worldminded and less ethnocentric, replies from 13 of the 17 original Ss in 1955 showed that their sharp decrease in ethnocentrism following the 1950 experience was still evident but that they had reverted to the relatively nationalistic attitude they held prior to the summer's experience, according to a significant decrease in W scale scores between October, 1950, and May, 1955 ($p < .005$, $t = 3.81$).

Attitude Stability and Interim Activities

In his study of volunteer work camps in the United States, Riecken (2) found a close relationship between a person's attitudes subsequent to the experience and certain of his activities after the experience. From a comparison of the attitudes of Ss who had maintained varying degrees of contact with one another and with the American Friends Service Committee during a 10-month period following the work-camp experience, he concluded that "continued reenforcement of new attitudes through sustained postcamp contact with like-minded individuals is the best single determinant of attitude stability that we are aware of" (2, p. 137).

Replies to Riecken's questions by Ss in the present study were coded according to his criteria for determining intensive, average, and little contact. The Experimenters constituted the most appropriate sample with which to test his hypothesis in view of the sizable number of Ss and of the special emphasis placed by the Experiment organization on the development of worldminded and nonethnocentric attitudes. There were 12 Ss (16 per cent of the Experiment sample) who had maintained intensive contact during the four-and-a-half year interval (marriage or close friendship with persons met through the Experiment and/or project or organizational work for the Experiment) and 18 (24 per cent of the Experiment sample) who had had little contact (active correspondence only or no contact other than receipt of Experiment literature).

An appropriate test of Riecken's hypothesis (although not the test that he employed) is to compare test-retest correlation coefficients of the intensive-contact and little-contact samples on the W and E scales. By the hypothesis of greater attitude stability for the intensive-contact sample, (a) the long-run coefficients for intensive-contact Ss (based on the October, 1950, and May, 1955, testings) should be significantly higher than the comparable coefficients for little-contact Ss, and (b) the long-run coefficients compared with the short-run coefficients (based on the May, 1950, and October, 1950, testings) should reveal no significant change within the intensive-contact sample, while within the little-contact sample the long-run coefficients should be significantly lower. With respect to the first prediction, a test of the significance of the difference between the r s for the two samples, using Fisher's z , yields no values beyond chance expectancy, although on both attitude scales z is considerably higher for the May-October, 1950, comparisons between the two samples—i.e., in contradiction to the implications of the hypothesis there is less difference in attitude stability between the two groups over the longer time-interval. Some support for the second pre-

diction appears in a significant decrease within the little-contact sample in the long-run correlation, compared to the short-run correlation, on the W scale ($p < .05$, $t = 2.33$ based on Hotelling's formula (8)); there are no other significant within-group changes.⁶

If Riecken's procedure is followed and the W- and E-scale means for the Ss in the two groups are compared for the three testing periods, the hypothesis is not supported. As the means in Table 2 reveal, there is much the same degree of difference between intensive-contact and little-contact Ss at all three testing intervals, with t being at chance expectancy in all cases. Furthermore, there is much the same trend in W- and E-scale means within both groups for the three testing periods.

These data do not support the hypothesis that attitude stability following the intercultural experience depends upon contact with likeminded individuals. The mean scores, in fact, suggest that the person who maintains contact for a long time after the experience does so because of an initially more worldminded, less ethnocentric orientation.

SUMMARY

Four and a half years after their return from a summer in Europe, a mail questionnaire was sent to 205 persons in four different experimental groups and to 10 Ss in a stay-at-home control group. Of these Ss, who had been tested initially both in the spring and in the fall of 1950, 156 returned questionnaires. The principal findings of the follow-up study are as follows:

1. Behavioral expression of cross-cultural friendship, as measured by correspondence and the sending of gifts over a long time-span following the intercultural experience, is most likely to transfer from a setting that facilitates the development of intensive personal relationships between Americans and foreigners. On the whole, both the short-run and long-run behavioral effects of the experience are limited in nature.

2. Although the original study revealed no short-run changes in general social attitudes, during the subsequent four-and-a-half year period Ss became significantly less worldminded, less ethnocentric, less authoritarian, and more conservative in their political and economic views. Events that have occurred since the summer in Europe appear to have had a far greater impact on the social attitudes of the Ss than the intercultural experience itself.

3. The data do not support a hypothesis that attitude stability following the intercultural experience depends upon contact with likeminded individuals. Evidence does suggest that the person

⁶ The eight product-moment r 's ranged from .66 to .895.

who maintains contact for a long time after the experience does so because of an initially more worldminded, less ethnocentric orientation than the person who loses contact.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. RIECKEN, H. W., JR. *The volunteer work camp: A psychological evaluation*. Cambridge: Addison-Wesley, 1952.
3. RIEGEL, O. W. Residual effects of exchange-of-persons. *Publ. Opin. Quart.*, 1953, **17**, 319-327.
4. SAMPSON, D. L., & SMITH, H. P. A scale to measure worldminded attitudes. *J. soc. Psychol.*, in press.
5. SMITH, H. P. Changes in attitude resulting from experiences in foreign countries. Unpublished doctor's dissertation, Harvard Univer., 1954.
6. SMITH, H. P. Do intercultural experiences affect attitudes? *J. abnorm. soc. Psychol.*, 1955, **51**, 469-477.
7. SMITH, M. B. A perspective for further research on cross-cultural education. *J. soc. Issues*, 1956, **12**, 56-68.
8. WALKER, H., & LEV, J. *Statistical inference*. New York: Holt, 1953.
9. WATSON, JEANNE, & LIPPITT, R. *Learning across cultures: A study of Germans visiting America*. Ann Arbor: Institute for Social Research, Univer. of Michigan, 1955.

Received April 9, 1956.

LEADERSHIP AND THE SPATIAL FACTOR IN SMALL GROUPS

GORDON HEARN¹

University of Toronto

DURING discussion, people tend to interact more with those sitting opposite them in a circle than with their neighbors. This was the hypothesis proposed and supported by the results of a study reported by Bernard Steinzor some years ago in this Journal (5). In order to test the hypothesis, he had observed the meetings of two 10-member discussion groups, keeping a record of who followed whom in the interaction sequence.

In the course of an experiment (3) conducted for a quite different purpose, data were collected which permitted a replication of Steinzor's study under fairly similar conditions and an extension of his hypothesis to include a consideration of the influence of leadership upon the spatial factor.

METHOD

The subjects (Ss) in the present study were 30 male students enrolled in an introductory psychology course. For eight weeks of the course they were divided into six six-member groups, the investigator, as leader, constituting the sixth member in all of them. The purpose of the experiment was to test the relative effectiveness of two training methods in improving group productivity. Accordingly, three groups were assigned to what was called the *Self-Motivated Method* while the three remaining groups experienced the *Trainer-Induced Method*. These two methods differed essentially in the behavior of the trainer or leader and the assumptions about training upon which his behavior was based.

Training in the self-motivated method was based on

an assumption that the group itself had the capacity and resourcefulness to improve its own functioning. Furthermore, it was assumed that the group, as the object of the training, should be actively involved in the direction of the training process. The leader was present and available should the group desire his help but he did not impose himself on the group. Throughout, his principal objective was to encourage the group to take maximum responsibility for improving its own performance. Thus, this method was somewhat similar to, although probably even less directive than, the Rogerian type of group-centered leadership used in Steinzor's Group I (4).

The trainer-induced method was based upon the assumption that an outside agent, in the person of a trainer or leader, is required to enable a group fully to understand and modify its behavior. He is an expert who knows better how the group needs to change in order to improve its performance. As such, in this method, the leader was very active during training, giving critical analyses of the group members' past performance, and functioning somewhat as a "coach" in urging them to modify, in certain ways, their future group behavior. Thus, in this case, the leader assumed major responsibility for improving the group's performance.

In the course of the experiment, the groups participated in eight two-and-a-half-hour weekly sessions. A typical session began with a training period which attempted to identify the ways in which the group needed to change in order to do better in the future, and each session concluded with a trial period in which the group was given a human relations problem to discuss and the task of formulating, as a group, a written solution to the problem. The members of each group and the leader, seated as shown in Fig. 1, occupied the same seats for all eight sessions. All groups had different problems from week to week but the same problem in any given week.

¹ The author, on leave from the University of California, Berkeley, 1955-56, acknowledges his debt to the Institute of Social Sciences and to the Cassidy Memorial Research Fund of the University of Toronto for material assistance in the preparation of this report.

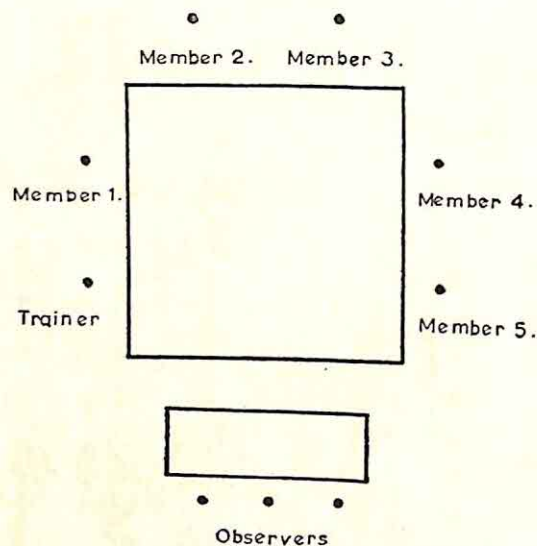


FIG. 1. SEATING ARRANGEMENT DURING EXPERIMENT

Using the Bales categories (1), an observer coded the interactions which occurred during the training periods. The "initiator," "object," and "interaction type" of each comment were coded for Sessions 4, 5, 6, 7, and 8, providing the data with which the Steinzor hypothesis was tested. Where Steinzor had recorded *who followed whom* in an interaction sequence, the present study noted *to whom* a comment was directed.

These two measures appear to be sufficiently related, however, to justify a comparison of the results of the two studies. Except for the fact that Steinzor's groups sat in a circle whereas Ss in the present experiment sat in a U-shaped arrangement, and that his groups were somewhat larger, the groups in the two studies operated under essentially similar conditions.

Steinzor's method of data analysis was employed in the present study. In a six-member group a comment can be directed to any one of the following positions: one seat to the left—1L, two seats to the left—2L, three seats away—3, two seats to the right—2R, or one seat to the right—1R. It should be noted, however, that in the U-shaped arrangement, interaction with one person to the left or to the right (1L or 1R) of the trainer or Member 5 could be either an interaction with a person across the table or an interaction with one's neighbor. This is a shortcoming of the present design which clouds the results to some extent. In the present study, the data from the self-motivated groups, I, II, and III, were combined as were those from the trainer-induced groups, IV, V, and VI, but were treated separately for Sessions 4, 5, 6, 7, and 8, these being the only weeks in which the essential data were collected. The results are shown in Fig. 2. There were instances in which a comment was coded as being directed to the group-as-a-whole. All such were excluded from the analysis. Analysis included all comments made by and toward the leader as well as those of the members. Table 1 shows for both studies those positions in which the obtained varied from the expected frequencies by an amount greater than expected at the .02 level of confi-

TABLE 1
DISTANCES WITH FREQUENCIES YIELDING CHI-SQUARE VALUES SIGNIFICANT AT AT LEAST THE .02 LEVEL OF CONFIDENCE

a. Leaderless (Steinzor)										
Group Session	1	2	3	4	5					
No. of positions possible	9	7	8	7	9					
Overchosen positions*	5	4	—	4	4R					
Underchosen positions	2L	—	—	—	1L, 1R					
b. Self-Motivated (Hearn)										
Group Session	4	5	6	7	8					
No. of positions possible	5	5	5	5	5					
Overchosen positions	2L, 3, 2R	2L, 3	3	—	3					
Underchosen positions	1L, 1R	1L, 1R	1R	2L	1L, 1R					
c. Group-Centered (Steinzor)										
Group Session	1	2	3	4	6	7	9	12	16	18
No. of positions possible	9	7	9	9	9	9	9	9	8	7
Overchosen positions	4R	2L, 2R	5	—	3R	—	—	5, 4L	—	2L
Underchosen positions	—	1R	3L	—	1L	—	—	2L, 2R	1R	—
d. Trainer-Induced (Hearn)										
Group Session	4	5	6	7	8					
No. of positions possible	5	5	5	5	5					
Overchosen positions	—	1L, 1R	—	—	—					
Underchosen positions	—	2L, 3	1L, 1R	2L, 3	—					

* The "overchosen position" row shows the session expected frequency. The "underchosen positions" row shows the session expected frequency.

* The "overchosen position" row shows the seating positions in which the observed frequency was significantly greater than the expected frequency. The "underchosen position" row indicates the reverse, where the observed was significantly less than the expected frequency. Thus in session 1 of Steinzor's leaderless group, the frequency of comments over a 5-seat interval significantly exceeded the expected, while over the two-seat-to-the-left interval frequency of comments was significantly less than the expected.

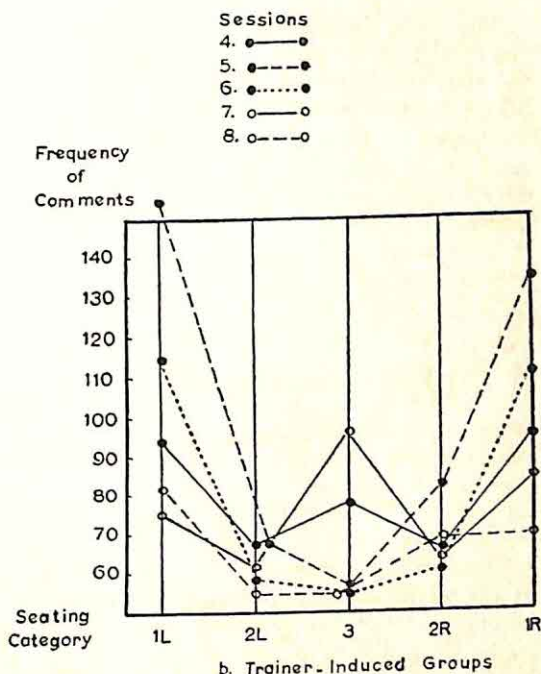
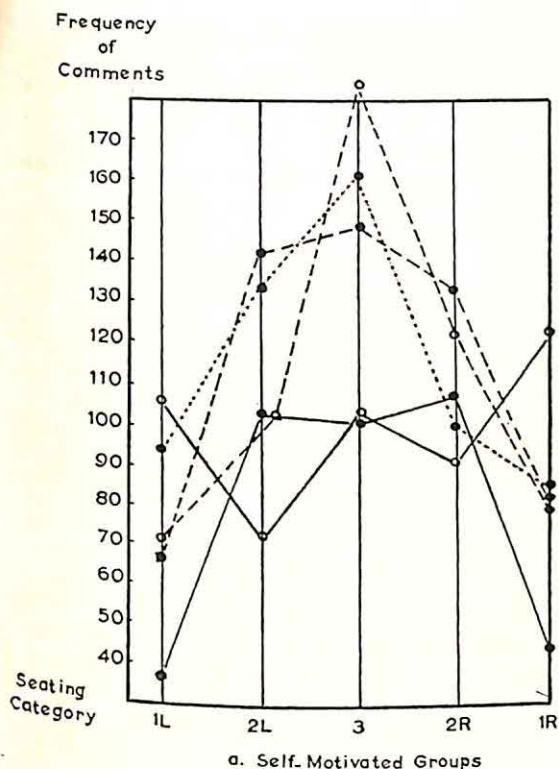


FIG. 2. DISTRIBUTION OF COMMENTS BY SEATING CATEGORY

dence,² according to chi-square analyses. The expected frequencies were derived from the assumption that the comments would be distributed equally into the five seating categories.

RESULTS

Data with respect to the self-motivated groups, as shown in Fig. 2a, give strong support to Steinzor's hypothesis. With the exception of Session 7, interaction is greater at the two- and three-seat intervals than at the one-seat interval. When the data from the five sessions are combined, the distribution into the five seating positions, 1L, 2L, 3, 2R, and 1R, deviates from the expected by an amount that permits us to reject the null hypothesis at the .001 level of confidence. In the self-motivated groups, therefore, there is strong evidence that people tend to interact more with persons sitting at a greater distance but more in their line of vision than with their neighbors to either side.

The situation in the trainer-induced groups, however, is just the opposite. An examination of Fig. 2b will show, again with the exception of Session 7, that there are more interactions at the one-seat interval than at the two- or three-seat

intervals. Again, the combined distribution into the five positional categories deviates from the expected by an amount which is significant at the .001 level of confidence. There is strong evidence in the trainer-induced groups, then, that members tend to interact more with their neighbors on either side than with those at a greater distance. The atypical results in both group types in Session 7 remain unexplained.

Since the groups had been assigned randomly to their training method, it seems valid to assume that the principal difference between the self-motivated and trainer-induced groups was the manner in which they were led. The active, directive kind of leadership in the trainer-induced groups apparently produced the reverse of the effect noted in Steinzor's study.

In order to analyze the possible interactive effect of leadership and the spatial factor, the data from the two studies are presented as is shown in Table 1. And in order to facilitate comparison, they are presented approximately in the form used in the Steinzor article. They are ordered according to the amount of leader direction assumed to be inherent in each group type. Accordingly, because of its absence of designated leadership, Steinzor's "Leaderless" group is listed first. Hearn's "Self-Motivated," in which the leader interjected himself only when drawn in by the group, is second. Steinzor's "Group-Centered" is listed third, for

² The .02 level was selected in the present study in order to facilitate comparison with Steinzor's study which had used the .02 in preference to the more conventional .01 or .05 criteria.

while it was quite similar to the "Self-Motivated," it probably involved somewhat more activity on the part of the leader. Admittedly this was an arbitrary decision, but it seemed justified, nevertheless, on the basis of the descriptions of the two methods. Hearn's "Trainer-Induced" method which represented the highest degree of leader direction is listed last.

When the four methods were placed in this sequence, one observed an interesting trend. Steinzor's hypothesis was most strongly supported in the "Leaderless" and "Self-Motivated" groups. In the case of Steinzor's "Group-Centered," however, the effect began to break down. Those seating positions which were overchosen to a significant degree, in one session or another, covered a range from the two- to the five-seat interval while those underchosen to a significant degree ranged from one to three. Thus, there was an overlap and the results were somewhat equivocal. When one looked at the "Trainer-Induced" groups, the effect had become reversed. Greater activity by a designated leader seemed to have the effect of reducing the distance, in terms of seating interval, over which members would direct their person-to-person comments.

DISCUSSION³

Thus the present study appears to support as well as extend the Steinzor hypothesis. However, because it was conducted under somewhat different conditions of size and seating arrangement, it cannot be regarded as a definitive test. Furthermore, because the ordering in Table 1 of the four leadership methods in the two studies was made on the basis of assumption rather than established fact, these data do not warrant firm conclusions concerning the interactive effect of leadership and the spatial factor. They do, however, suggest the following hypotheses:

1. When direction by a designated leader is at a minimum, members of a face-to-face discussion group will direct more comments to those sitting opposite them than to their neighbors on either side.
2. When direction by a designated leader is at a maximum, members of a face-to-face discussion group will direct more comments to their neighbors on either side than to those sitting opposite them.

³ The author is indebted to Leon Festinger, Thomas Gordon, Henry S. Maas, Carl R. Rogers, and Bernard Steinzor for a number of helpful suggestions in the interpretation of the results of this study.

3. When direction of the face-to-face discussion group is shared about equally by the members and the designated leader, factors other than the spatial factor will determine which will be the underchosen and overchosen seating intervals.

The present study suggests, further, that other group situations which, when observed in the future, reveal "no spatial influence," may not actually be negative findings but rather cases in which other conditions are in play to cancel out or overpower the spatial factor. One such case in point is a study by Bass and Klubeck (2) which suggests that the leadership status of group members in situations outside the experiment may operate to diminish and almost eliminate the Steinzor effect. And undoubtedly other factors which interact with the spatial factor may also be found.

SUMMARY

In an earlier study, Steinzor showed evidence to support the hypothesis that persons tend to interact more in group discussion with persons sitting opposite them than they do with their neighbors. The present study, conducted under somewhat different conditions, supports but also extends this hypothesis. Its results, in combination with those of Steinzor, suggest a further hypothesis to the effect that the manifestation of the *Steinzor effect* will depend upon the degree of direction given by a designated leader. That is, the effect will be most strongly manifested in the groups with no designated leader. It will tend to disappear in groups where group direction is shared about equally by the members and the designated leader, and it will be reversed in situations where the designated leader gives very strong direction.

REFERENCES

1. BALES, R. F. *Interaction process analysis*. Cambridge, Mass.: Addison-Wesley, 1950.
2. BASS, B. M., & KLUBECK, S. Effects of seating arrangement on leaderless group discussions. *J. abnorm. soc. Psychol.*, 1952, **47**, 724-727.
3. HEARN, G. The training of discussion groups—An experimental study. Unpublished doctor's dissertation, Massachusetts Institute of Technology, 1948.
4. STEINZOR, B. The development and evaluation of a measure of social interaction. *Hum. Relat.*, 1949, **2**, 319-347.
5. STEINZOR, B. The spatial factor in face to face discussion groups. *J. abnorm. soc. Psychol.*, 1950, **46**, 552-555.

Received April 13, 1956.

JUDGMENTS OF LEADERSHIP BASED UPON PHYSIOGNOMIC CUES¹

DONALD J. MASON²

University of Illinois

THE existence of physiognomic stereotypes for a number of personality characteristics has been firmly established (1, 2, 4, 9, 10). In short, it appears that people will with some consistency, whether correctly or incorrectly, attribute certain personality characteristics to individuals possessing certain physiognomies. It has been hypothesized that such stereotypes are of importance in determining how individuals react toward one another in social situations. Krech and Crutchfield, for example, report that "... imposing physical stature, a 'strong' face, ... an air of self-confidence and competence all predispose the group to accept an individual as leader" (8, p. 437). In the area of the psychology of leadership there is also abundant evidence which indicates that there is a low positive correlation between certain physical characteristics (e.g., height, weight, and stamina) and temperament, peer status, and achievement. Gowan (5), for example, reports that leaders in many professions tend to be taller on the average than their subordinates. Other studies (3, 6, 11) indicate that individuals with athletic physiques tend to be more dominant and self-assertive than those who lack such physiques.

The present study represents an attempt to determine whether naive subjects can judge the relative leadership abilities of individuals from the cues provided by photographs of the individuals.

METHOD

Criterion of leadership. The external criterion for leadership for this study consisted of the intragroup rankings assigned to 75 candidates for the Colorado State Highway Patrol on the basis of their performances in a field test in which 16 different small groups of from four to six candidates were presented with the problem of locating and effecting the capture of a group of armed desperadoes and the release of a hostage. The situation was structured in such a manner as to require a maximum of cooperation from the candidates in order to achieve a satisfactory solution to the problem. The candidates were rated by three skilled judges on the basis of a semiobjective check list of characteristics of leadership manifested by the candidates in the test situation and in a discussion period which followed. Subsequent to the field test and discus-

sion period the candidates were also required to rank the members of their group in terms of who was the best leader, the second best, and so on. The high degree of interjudge agreement among the three expert judges is indicated by a coefficient of concordance (7) of .91. The mean coefficient of concordance between the mean ranking assigned each candidate by the judges and the mean rank assigned by his peers was .78. These findings are taken as indicative of some degree of reliability and construct validity of the criterion measure.

Procedure. The Ss used in the present study were 24 undergraduate college students enrolled in an introductory psychology course. The individual Ss were presented with the groups of small individual photographs of the candidates and were instructed to rank the four to six photographs in each group in terms of "which person would be the best leader, the second best, and so on." The same Ss ranked both full-figure photos and head-and-shoulders photos in this manner. The order of presentation of the different groups of photos and the various individual photos within the groups were randomized.

RESULTS AND DISCUSSION

The mean rank difference correlation between the rankings assigned by the Ss on the basis of the head-and-shoulders photos and the criterion was .18 ($SD = .5$). The mean rank difference correlation between the rankings assigned by the Ss on the basis of the full-figure photos and the criterion was .04. These correlations fall far short of statistical significance and indicate that the Ss were not able to judge leadership, as here defined, from photographs of the individuals.

The mean rank difference correlation between the rankings assigned by the Ss on the bases of the two different sets of photos (head-and-shoulder versus full-figure) was .80. This correlation suggests that the same cues were used by the judges for ranking both sets of photos.

Coefficients of concordance were computed to provide an estimate of the degree to which the individual Ss tended to agree with one another in their judgments of the photographs. This procedure yields, in essence, a measure of the reliability of the judgments as opposed to the validity of the judgments (indicated by the correlations with the criterion). The mean coefficient of concordance for the 16 groups of full-figure photos was .30. Of the 16 coefficients of concordance for the various groups, 14 of these were statistically significant at or beyond the .05 level of confidence and the other two approached significance. This finding supports the hypothesis that the Ss share a common conception of what a leader looks like,

¹ The data upon which this study is based were gathered while the author was at the University of Colorado. The author wishes to acknowledge his gratitude to Dr. K. R. Hammond who made material contributions to all phases of this research.

² Public Health Service Research Fellow of the National Institute of Mental Health.

even though this physiognomic stereotype does not correlate with the criterion measure for leadership.

REFERENCES

1. ALLPORT, G. W. *Personality: a psychological interpretation*. New York: Holt, 1937.
2. ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258-290.
3. BAYLEY, NANCY. Some psychological correlates of somatic androgyny. *Child Developm.*, 1951, **22**, 47-60.
4. BRUNSWIK, E., & REITER, L. Eindruckscharaktere schematisierter Gesichter. *Z. Psychol.*, 1937, **142**, 67-134.
5. GOWAN, E. B. *The executive and his control of men*. New York: MacMillan, 1915.
6. HANLEY, C. Physique and reputation of junior high school boys. *Child Developm.*, 1951, **22**, 247-260.
7. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
8. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
9. SECORD, P. F., BEVAN, W., & DUKES, W. F. Occupational and physiognomic stereotypes in the perception of photographs. *J. soc. Psychol.*, 1953, **37**, 261-270.
10. SECORD, P. F., DUKES, W. F., & BEVAN, W. Personality in faces: I. An experiment in social perceiving. *Genet. Psychol. Monogr.*, 1954, **49**, 231-279.
11. SHELDON, W. H. *The varieties of temperament*. New York: Harper, 1942.

Received April 19, 1956.

THE EFFECTS OF ANXIETY AND OF SITUATIONAL STRESS ON COMMUNICATIVE EFFICIENCY¹

RUTH AUTREY GYNTHNER

University of North Carolina

ALTHOUGH most studies of the relationship between anxiety and performance have been concerned with learning tasks (6, 7, 10, 11), there has been some interest in the effects of anxiety on speech behavior (1, 2). Since stress and anxiety tend to foster disorganization, the language of the individual under these conditions would be expected to show disruption of the organizing process as evidenced by fragmentary speech, overinclusion, and other signs of disorganization (5). Thus, the efficiency of the communication is lowered. Also serving to lower the communicative efficiency would be verbal responses elicited by the anxiety itself and the increased relative strength of incorrect responses under conditions of high drive. These considerations lead to the hypothesis that under conditions of stress and anxiety an individual's communicative efficiency is lowered.

METHOD

Subjects for this study were selected from a group of 307 students in elementary psychology to whom the Welsh A scale (15) had been administered. This scale consists of 39 items selected by factor-analytic methods from the Minnesota Multiphasic Personality Inventory.

¹ This paper is based on a Master's thesis submitted to the graduate school of the University of North Carolina under the author's maiden name, Olive Ruth Autrey. Appreciation is expressed to Dr. G. S. Welsh, Dr. J. W. Thibaut, and Dr. W. G. Dahlstrom, members of the thesis committee.

It has been shown to be correlated with anxiety in both normal and psychiatric groups.

The range of scores in this college population was from 0 to 35 with a mean of 14.2. A group of 18 high-A or "anxious" Ss were selected from the upper 12 per cent of the range (scores from 24 to 35). The "non-anxious" or low-A group fell in the lower 15 per cent (scores 0 to 6).

These high- and low-A groups were each randomly divided into two subgroups of nine Ss each, termed the stress and nonstress subgroups. The stress subgroup was given the following introduction to the interview situation:

You have probably taken personality tests of the paper and pencil type and maybe some of the projective tests such as the Inkblot test. Well, we have here a technique which gives an even deeper understanding of an individual and his personality. This is obtained from the way he answers a few simple questions. Previous studies with this technique have indicated that it is the best means we now have for predicting a person's success in college and in his vocation. Not only is it useful in predicting success in achievement, it is quite revealing of the individual's self-concept and of underlying motives in his relationships with other people.

So here is the list of questions which you may answer in any way that you like. What you say will be recorded by this microphone. Now what is your name?

The following instructions given to the low-stress group were designed to de-emphasize the importance of the interview:

We are making a study of language and want a

number of samples of people talking, so just answer these questions any way that you like. Your name will not be connected with the interview in any way. We merely number the interviews and these numbers do not correspond with names. This mike will be used for recording. (Speaking into mike) This is interview number —.

After the introduction, each *S* was asked a list of 11 questions about himself. These questions were designed to prevent one-word answers and were later described by *Ss* as "vague," "ambiguous," and "hard to answer." Each *S's* responses were tape recorded and transcribed verbatim. The quantitative measure of communicative efficiency used was one devised by Calvert (3). In using this method, *Ss'* verbatim responses were divided into thought units and then studied from the standpoint of how well each thought met the minimum requirements of communication. Each thought unit was characterized as being either directly relevant, indirectly relevant, or irrelevant to the question asked. If the thought unit was directly relevant it was considered to have met the minimum requirements for communication and was classified as Basic Thought (BT). If it was indirectly related and served mainly to modify the BT it was classified as Secondary Material (SM). Together these two classes constituted the Communicative Material (CM). The ratio of the effective elements (the CM) in the verbalization to the total number of elements used was defined as the index of the individual's Communicative Efficiency (CE). Both letter and word counts were used in computing the ratio, the two units yielding almost identical results.

To demonstrate interscorer agreement, Calvert had each of three scorers analyze five interviews for BT and SM. The product moment correlations between the three judges ranged from .79 to .88 for BT and from .75 to .86 for SM. These correlations indicate an interscorer agreement beyond the .01 level of confidence.

Calvert demonstrated the consistency of the verbal behavior by readministering the interview to 12 *Ss* one and one-half to four months after the first interview. For Communicative Efficiency he found a correlation of .68, which is significant beyond the .01 level.

RESULTS

As shown in Table 1, the group means occurred in the predicted relationship. A chi-square test between the extreme within-cell variances was insignificant, so homogeneity of variance could be assumed with no further test. Table 2 shows that there was no significant interaction between stress and anxiety, and it was therefore possible

TABLE 1
SCORES ON COMMUNICATIVE EFFICIENCY (CE)
FOR THE FOUR GROUPS

Anxiety	Stress		Nonstress	
	Mean	SD	Mean	SD
Low	.508	.12	.647	.14
High	.433	.10	.550	.13

TABLE 2
ANALYSIS OF VARIANCE OF THE CE SCORES
WITH RESPECT TO ANXIETY (A) AND
STRESS

Source	Sum of Squares	df	F	t	p*
Between A	.067	1	3.94	1.98	.025
Between stress	.147	1	8.65	2.97	.005
Interaction	.002	1	—		
Error	.551	32			
Total	.767	35			

* Value for a one-tailed test.

to deal with the data as two groups of 18 cases rather than as four separate groups of nine cases each. Since the direction of results had been predicted, one-tailed significance values were obtained. Table 2 shows that the effect of anxiety on communicative efficiency was significant at the .025 level and that the effect of situational stress on CE was significant at the .005 level.

DISCUSSION

The results of this study indicate that people who score high on the Welsh A scale have lower CE scores than do people who make low scores on this scale. This finding is consistent with the theoretical expectation that highly anxious *Ss* have a less adequate performance level on complex tasks than do those with less anxiety. The experiment also shows that situational stress serves to lower the CE of both high and low anxiety *Ss* and that this stress appears to have the same effect on both groups. Since, in contrast with the present results, interaction between degree of anxiety and stress has been found in experiments on stress in a learning situation (e.g., 6), the equivalent effects reported here of stress on performance of both high- and low-anxiety groups may be a function of the type of task, i.e., communicating in a face-to-face situation in response to questions that are vague and personal in nature.

This experiment demonstrates a phenomenon that is repeatedly observed in everyday contacts between people. Impressions and evaluations of a person, whether in a business, social, or professional setting, are influenced by one's ability to communicate one's ideas. This study shows that the lack of fluency in an interview situation is in part a function of the individual's degree of anxiety and the amount of stress in the situation.

SUMMARY

To test the effects of experimental stress and of degree of anxiety on communicative efficiency two groups of 18 *Ss* each were chosen from students making extreme scores on the Welsh A scale.

Half of the Ss in each of these groups were given an introduction to the interview situation that was designed to produce stress and ego-involvement. The other half received an introduction that de-emphasized the importance of the interview. The Ss were then given a list of 11 questions and their responses tape recorded. The verbatim responses were analyzed according to Calvert's manual, and communication efficiency (CE) scores were obtained. An analysis of variance showed no interaction between anxiety and stress. The Ss low in anxiety were found to have CE scores that were significantly higher than those of Ss high in anxiety and stressed Ss to have lower CE scores than the nonstressed Ss. These results were interpreted as indicating that anxiety and stress interfere with a person's effectiveness of communication.

REFERENCES

1. BALKEN, EVA RUTH, & MASSERMAN, J. H. The language of phantasy: III. The language of phantasies of patients with conversion hysteria, anxiety state, and obsessive-compulsive neurosis. *J. Psychol.*, 1940, **10**, 75-86.
2. BENTON, A. L., HARTMAN, C. H., & SARASON, I. G. Some relations between speech behavior and anxiety level. *J. abnorm. soc. Psychol.*, 1955, **51**, 295-297.
3. CALVERT, JAMES J. Verbal behavior as a predictor of personality syndromes. Unpublished doctor's dissertation, Ohio State Univ., 1950.
4. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
5. HASSOL, L., CAMERON, N., & MAGARET, ANN. Production of scattered speech via personalized distraction: An investigation of continuity in normal and schizophrenic language. *Amer. Psychologist*, 1952, **7**, 351-352. (Abstract)
6. LUCAS, J. D. The interactive effects of anxiety, failure, and intra-serial duplication. *Amer. J. Psychol.*, 1952, **65**, 59-66.
7. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166-173.
8. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89-101.
9. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553-565.
10. RAMOND, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120-124.
11. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116-119.
12. SPENCE, K. W., & TAYLOR, JANET. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.
13. TAYLOR, JANET. The relationship of anxiety to conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
14. TAYLOR, JANET, & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.
15. WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univ. of Minnesota Press, 1956.

Received May 9, 1956.

MATURATION AND THE SPIRAL AFTEREFFECT

GEORGE F. HARDING, SIDNEY M. GLASSMAN, AND WILLIAM C. HELZ

United States Army

PRICE and Deabler (6) found a high correlation between failure to perceive the spiral aftereffect and the presence of organic brain pathology. Their conclusions led the present authors to speculate upon the maturational aspects of the phenomenon. Weinbach (8) reports that full brain weight is not developed until about 7 years of age and that electroencephalographic tracings show a developmental pattern through years eight to ten. Further, other experimenters (1, 2, 5) have established that many psychological phenomena first occur about this time. These studies suggest that a similar, though not necessarily identical, limitation of function exists in neurologically immature children and adults with organic cerebral damage. The specific hypothesis of this

experiment is that children below the age of cerebral maturation (6 to 8 years) will fail to perceive the spiral aftereffect.

SUBJECTS AND PROCEDURE

The Ss used were 81 children ranging from 48 to 71 months of age. No children with known organic brain pathology were used.

A 78 RPM phonograph turntable was used to rotate the Archimedes spiral. The turntable was placed on its side so that the Ss could stand facing it from a distance of approximately eight feet. The spiral was constructed of a white disc six inches in diameter upon which a black spiral of 920° was drawn. A second reversed spiral was constructed for the effect of opposite rotation.

The procedure followed was divergent from that of Price and Deabler in several respects. Upon entering

TABLE 1

DISTRIBUTION OF CHRONOLOGICAL AND MENTAL AGES AND SUCCESS OR FAILURE IN PERCEIVING THE SPIRAL AFTEREFFECT

CA (months)	N	Passed	Failed	MA (months)	N	Passed	Failed
46-48	2		2				
49-51	3		3				
52-54	3		3	52-54	1		1
55-57	9	2	7	55-57	1		1
58-60	15	8	7	58-60	2		2
61-63	8	4	4	61-63	5	1	4
64-66	13	6	7	64-66	2	1	1
67-69	7	3	4	67-69	6	5	1
70-72	5	5		70-72	7	5	2
73-75	3	3		73-75	2	2	
76-78	6	6		76-78			
79-81	2	2		79-81	2	2	
82-84	5	5		82-84	1	1	
Totals	81	44	37		29	17	12

the room, the Ss were shown the apparatus and a brief attempt was made to gain rapport. To insure that the Ss could conceptualize relative size, and to induce a set to look for size differences, one series of circles and another of lines of unequal size were shown them. They were asked to identify the smallest, then the largest, in each series.

The discs were rotated for a period of 30 seconds. During this time repeated instructions were given to "Keep looking at the center of the wheel. Don't look away." Just prior to stopping the disc the examiner said, "I'm going to stop it now. Don't look away." Then, "What do you see . . . ? What is the line doing now?" Disc A, resulting in an aftereffect of expansion, was shown first to half the Ss. Disc B (contraction) was shown first to the remainder.

A total of four trials in the order ABAB or BABA was given to each of the Ss. If it was necessary to clarify communication after the first or second trial, the question became, "Does it get bigger or smaller?" In instances where the Ss followed suggestion and guessed consistently, but incorrectly on the remainder of the trials, no credit was given. The Ss "passed" the test when they gave correct responses to at least three of the four trials. The Ss who failed to perceive the aftereffect were retested after an interval of about two weeks. This was done in order to evaluate whether failure was due to lack of rapport, inadequate communication, or genuine inability to perceive the effect. To explore the relationship of the phenomenon to mental age, a stratified sample of the Ss were administered the short form of the Stanford-Binet Intelligence Scale, Form L.

Every effort was made to determine whether or not these young Ss who did not report the aftereffect were truly not perceiving it or were merely unable to communicate the experience. The reaction of amazement and the assuredness of those who did report it convinced the experimenters of the validity of the verbal responses. Nevertheless, the "faulty communication hypothesis" cannot be completely refuted by this method.

RESULTS AND DISCUSSION

Table 1 presents the CA and MA distributions of these Ss in relation to success or failure in perceiving the spiral aftereffect. The youngest child to perceive the effect was 55 months of age; beyond the 69th month all children in the sample were successful. Insofar as failure to perceive the spiral aftereffect is a valid criterion of incomplete brain function, these results seem to indicate that the specific brain functions necessary to perceive the spiral aftereffect appears between 4.5 to 6 years of postnatal life.

The main differences between the MA and CA distributions is that there is less overlap of the Ss classified by MA and that several of the very young Ss who did perceive the aftereffect were found to be of considerably higher MA. This difference indicates that MA is a better criterion for predicting performance on the spiral than CA and possibly that more rapid cerebral maturation is present in the brighter Ss.

SUMMARY

The spiral aftereffect technique was administered to 81 children ranging in age from 48 to 71 months. No children below 55 months reported the aftereffect, while all of those above 69 months were able to report the phenomenon. Mental age was found to bear a more direct relationship to perception of the aftereffect than chronological age. The results of this study lend support to the hypothesis that children below a certain age level, presumably because of insufficient neural maturation, exhibit some behaviors similar to those of brain injured adults.

REFERENCES

1. BAYLEY, NANCY. Mental growth in young children. *Yearb. nat. Soc. Stud. Educ.*, 1940, **39** (II) 11-47.
2. BAYLEY, NANCY. Development and maturation. In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: Van Nostrand, 1951. Pp 145-199.
3. BRILL, N. Q., & SEIDEMANN, H. The electroencephalogram of normal children. *Amer. J. Psychol.*, 1941, **98**, 250-256.
4. LINDSLEY, D. B. Electroencephalography. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol 2. New York: Ronald, 1944.
5. PIAGET, J. *Play, dreams and imagination*. New York: Norton, 1951.
6. PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299-302.
7. STANDLEE, L. S. The Archimedes negative aftereffect as an indication of memory impairment. *J. consult. Psychol.*, 1953, **17**, 317.
8. WEINBACH, A. P. Some physiological phenomena fitted to growth equations. II Rate of growth of brain potentials (Alpha frequency) compared with rate of growth of brain. *Growth*, 1938, **2**, 247-251.

Received May 15, 1956.

THE PERCEPTION OF JEWISH PHOTOGRAPHS BY NON-JEWS AND JEWS

ALVIN SCODEL AND HARVEY AUSTRIAN

The Ohio State University

STUDIES by Allport and Kramer (2) and Lindzey and Rogolsky (6) have presented evidence that persons characterized by anti-Semitic attitudes have a greater tendency than persons of less prejudice to label photographs as Jewish. Moreover, their data indicate that the more prejudiced not only have such a response bias, but are actually more accurate in their identifications of Jews from photographs than are the less prejudiced. However, a subsequent study by Carter (4) found no consistent relationship between prejudice and accuracy of identification, and Elliott and Wittenberg (5) found that persons high on an anti-Semitism measure were more accurate only when the percentage of Jewish photographs in the sample was high. They explained this finding on the basis of the response bias to label photographs as Jewish in the anti-Semitic group.

The present study has two purposes. The first is to determine the relationships, in both a Jewish and non-Jewish group, between authoritarian attitudes and, on the one hand, response bias in labeling photographs as Jewish, and, on the other, the accuracy of identification of such photographs. The second purpose is to ascertain if there are differences in response bias and accuracy between the Jewish and non-Jewish groups.

METHOD

The non-Jewish and Jewish groups consisted of 194 and 118 Ss respectively. All non-Jewish Ss were selected from the introductory psychology class at Ohio State University. About one-third of the Jewish Ss were obtained from the same source. The remainder were recruited from the local Hillel Foundation and one of the campus Jewish social fraternities. In general, the groups were comparable with respect to age, sex distribution (about 35% females in each group) and educational level. Inasmuch as a direct measure of anti-Semitism would have little practicality for Jewish Ss, a measure was needed which would tap personality dispositions related to antiminority attitudes. The California F scale was chosen to fulfil this need since one of the ostensible purposes in its construction was to measure antiminority attitudes indirectly. The reported correlation between the F scale and Ethnocentrism (E) is .75 (1, p. 262) and that between the F scale and Anti-Semitism (A-S) is .53 (1, p. 263). Accordingly, each S was administered a 28-item F scale and was asked to write his religious affiliation on the face sheet (this procedure was obviously not necessary for the Hillel and fraternity Ss). The Ss who did not list religious affiliation were questioned further until a basis for dichotomy between Jewish and non-Jewish could be established.

Each S was individually shown a group of 100 photographs of adult males taken from popular magazines and pasted on cardboard posters. The choice of photographs was dictated by several considerations. Previous studies in this area have used yearbook photographs of college students which tend to have a highly uniform quality. The present writers felt that, short of having actual persons as stimuli, the best approximation to reality was to maximize available cues by adopting the more Brunswikian procedure of diversification in the photographs with respect to such variables as age, occupation, clothing, and facial expression. The second consideration involved the distributions of ethnic groupings and the number of Jewish photographs to be included. Since Carter's study demonstrated that Jewish photographs could not be identified better than chance when the non-Jewish photographs were of people of Mediterranean origin, the present sample of photographs was designed for moderate difficulty by including, in addition to Jewish photographs, those of various European groups as well as Americans. Thirty Jews (27 American and 3 European) were included in the sample of 100. Of the remaining 70, 30 were Americans, 20 were Northern Europeans (English, Scandinavian, and German) and 20 were from Latin countries (French, Spanish, and Italian). The 30 Jewish photographs included such disparate personalities as Arthur Koestler, General Grandval, David Selznick, and Harry (the Cap) Lev.¹ The unidentifiability of all 100 photographs was insured by pretesting with a larger sample of photographs with six graduate students as Ss. Recognized photographs were discarded. In the total sample of 312 Ss there were less than 20 identifications, and these appeared to be randomly distributed among the pictures as well as the two groups of Ss.

In the experimental procedure S first made a judgment of "Jewish" or "not Jewish" for each of the 100 photographs. Upon completion of this part, he was asked to go through the pictures again and check on a separate answer sheet the numbers (the photographs were numbered from 1 to 100) of the 30 photographs he thought were most likely to be photos of Jewish persons. The Elliott and Wittenberg study suggested that accuracy of perception of Jewish photographs is a function of the interaction between response bias and the percentage of Jews in the sample. The present procedure circumvents this difficulty by obtaining an accuracy score (ranging from a possible 0 to a possible 30) independent of response bias.

¹ Some ticklish problems of definition were deliberately ignored. For example, Arthur Koestler was included in the Jewish sample, although he no longer considers himself to be Jewish. For purposes of this study, the Jewish photos were defined in terms of the individual's ethnic origin, regardless of his self-perception.

TABLE 1

DIFFERENCES IN RESPONSE BIAS AND ACCURACY FOR HIGH AND LOW NON-JEWISH AND JEWISH F SCORES

Group	Range of F Scores	Response Bias			Accuracy		
		Mean number of Jews	SD	t	Mean number of Jews	SD	t
Non-Jewish Highs (N = 49)	123-163	39.35	13.71	2.99**	14.48	2.73	1.62
Non-Jewish Lows (N = 49)	33-85	30.73	14.64		13.53	2.55	
Jewish Highs (N = 30)	113-142	41.87	12.34	1.96*	14.93	2.22	.82
Jewish Lows (N = 30)	34-80	35.83	11.10		15.36	1.74	

* Significant at .05 level.

** Significant at .01 level.

TABLE 2

DIFFERENCES IN RESPONSE BIAS AND ACCURACY BETWEEN THE NON-JEWISH AND JEWISH GROUPS

Group	Response Bias			Accuracy		
	Mean	SD	CR	Mean	SD	CR
Non-Jewish (N = 194)	35.65	13.83	1.88	14.04	2.76	2.61**
Jewish (N = 118)	38.42	12.84		14.89	2.81	

** Significant at .01 level.

RESULTS

Upper and lower quartile F scores were selected within each group for comparisons of response bias and accuracy. These data are summarized in Table 1. For both the non-Jewish and Jewish groups there are significant differences between the extreme F-score groups in the number of photographs labeled Jewish. However, these data do not support the contention that the higher in prejudice (or, in this instance, authoritarianism) are more accurate in their identification of Jewish photos. There is no significant relationship, moreover, between response bias and accuracy. The correlations between response bias and accuracy of identification were $-.11$ for the total non-Jewish group and $+.08$ for the total Jewish group, neither of which was significant.

Comparisons between the non-Jewish and Jewish groups are summarized in Table 2. The Jewish group is more accurate than the non-Jewish group, and there is also a decided trend in the direction of a greater response bias in the Jewish group ($p = .06$). The differences in accuracy between the groups were actually more pronounced for low scorers than high scorers. A chi-square test between non-Jewish and Jewish low-quartile scorers resulted in a significant difference ($\chi^2 = 4.97$, $p < .05$), but no significant difference was obtained between the high scorers ($\chi^2 = .70$, $p > .40$). It should be noted that F scores for these comparisons were not equated because of the lower F scores in the Jewish group.

DISCUSSION

The present results support previous findings that persons with antiminority attitudes, inferred in this instance from F-scale scores, have a response set to label photographs "Jewish." In addition, Jews who are high on F reveal much the same response bias in contrast to those who are low. Whether high F scores in a Jewish population are related to anti-Semitism, i.e., self-hating attitudes, is a complex matter beyond the scope of this study. It does seem likely, though, that a minority group's adoption of selected hostile aspects of the dominant cultural ethic will lead to partial acceptance, at least, of that culture's stereotypes, and, for that reason, high scoring Jews would be expected to show the same response bias as high scoring non-Jews.

There is, however, the seeming paradox that the Jewish sample, as a group, manifests a tendency to see more of the photos as Jewish ($p = .06$) even though the F scores in the Jewish sample are significantly lower. The mean number of photos termed Jewish by the Jewish sample was 38.42, over eight more than were included in the 100 photographs. We would tentatively suggest that the vigilance hypothesis advanced by Lindzey and Rogolsky to account for the response bias and greater accuracy of anti-Semites is as applicable, if not more so, to Jews than anti-Semites. Our results are compatible with those of Elliott and Wittenberg in finding no relationship between antidemocratic attitudes and accuracy of identification of Jews. Nonetheless, Jews, as a group, tend to give more false positives and are more accurate than non-Jews. We would reason that many minority group members are particularly sensitive to possible cues in others which would enable them to ascribe same-group membership to these others and that this disposition is directly related to the degree of acceptance of the majority stereotype. Presumably, security is enhanced by this type of projective assimilation. Those with equalitarian attitudes are less likely to look for observable differences between non-Jews and Jews although past experience with both groups

will result in just as accurate judgments when they are forced to make selections. Parenthetically, the absolute difference in accuracy between non-Jews and Jews, although highly significant, is still quite small (less than one out of a possible range of 0-30). A greater difference would undoubtedly be expected if actual persons rather than photographs were used as stimuli, but it is also possible that, although Jews display more sensitivity, most of the cues entering into the perception of Jews are in the public domain. Over 50 per cent in both groups had 14 or more correct identifications. Application of probability functions of a hypergeometric distribution to the present sample resulted in a probability of .005 for 14 or more accurate identifications for any one individual so it is clear that both groups performed demonstrably better than chance in selecting Jewish photographs. Whether Jews can be identified better than chance, however, is a relatively meaningless question since accuracy is obviously dependent on the nature of the group from which Jews are being selected.

The relationship between F-scale scores and response bias in the non-Jewish group can best be explained on the basis of a passive acceptance of ethnic stereotypes in the high scoring group. Minimal resistance is offered to any cues which are usable for incorporation into the stereotype. A methodological problem arises in connection with the differences between high and low scorers in response bias since it might be argued that high Fs, extrapolating from Bass's findings (3), would manifest a greater response bias with respect to any variable. That is, the acquiescence of high Fs could result in a higher number of acceptances of any group, including Jews, but if such were the case, the specificity of the acceptance of ethnic stereotypes is seriously impugned. As part of another study, therefore, a comparable group of Ss who had been administered the F scale were exposed to the same set of pictures used in the present study and were asked to indicate, first, those photographs appearing to be those of businessmen and, second, those of men over the age of 50. For the latter categorization upper and lower quartile Fs had practically identical means, but in the categorization of businessmen the low scorers had a significantly higher number ($p < .05$). It can be concluded then that the tendency of high scorers to label more photographs as Jewish cannot be attributed to mere acquiescence.

One incidental finding deserves mention. Seven-

teen Ss in the two groups named 15 or less of the photos as Jewish even though their accuracy scores in the second part of the study were comparable to those of other Ss. Eleven of these Ss were low scorers and only two were high scorers. In view of the fact that there were 30 Jewish photographs, it can be hypothesized that some low scorers on measures of authoritarianism or prejudice display an exaggerated resistance to stereotypes. The wish to appear liberal may take the form of an extreme denial of ethnic differences even when such differences are apparent to the person. Certainly this subtype of low scorer merits further exploration.

SUMMARY AND CONCLUSIONS

Two groups of Ss, 118 Jewish and 194 non-Jewish, labeled a group of 100 photographs as Jewish or non-Jewish. In addition, each S selected 30 of the 100 that appeared to be the most Jewish-looking. The photographs included 30 Jews and, in the non-Jewish group, 30 Americans, 20 Northern Europeans and 20 Southern Europeans. Each subject was also administered the California F scale.

In both groups high F scorers labeled more photographs as Jewish than low F scorers, but in neither group was there a significant difference in accuracy between high and low scorers. There was no significant relationship in either group between response bias and accuracy. With respect to between-group differences Jews were more accurate than non-Jews and manifested a tendency to label more photographs as Jewish. Speculations were offered to account for the preceding results.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9-39.
3. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.
4. CARTER, L. F. The identification of "racial membership." *J. abnorm. soc. Psychol.*, 1948, **43**, 279-286.
5. ELLIOTT, D. N., & WITTENBERG, B. H. Accuracy of identification of Jewish and non-Jewish photographs. *J. abnorm. soc. Psychol.*, 1955, **51**, 339-341.
6. LINDZEY, G., & ROGOLSKY, S. Prejudice and identification of minority group membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 37-53.

Received May 18, 1956.

INTERRELATIONS AMONG THREE CRITERIA OF IMPROVEMENT IN GROUP THERAPY: COMFORT, EFFECTIVENESS, AND SELF-AWARENESS¹

HERBERT C. KELMAN AND MORRIS B. PARLOFF

National Institute of Mental Health, U. S. Department of Health, Education and Welfare

THE widespread use of global evaluations of improvement in psychotherapy assumes that improvement is a unitary phenomenon, i.e., that the different criteria of improvement vary together. Clinical evidence suggests that this assumption may be unwarranted. It becomes necessary, therefore, to spell out separate criteria of improvement and to develop measures for each. Only when specific criteria and objective measures are available will it be possible to gain some understanding of the nature of changes produced in therapy and the process by which such changes are effected.

Criteria of improvement have been proposed by Kogan and Hunt (6), who distinguish three points of view from which psychotherapy is evaluated: that of the person who receives the psychotherapy, that of the other persons who come in contact with him, and that of mankind more generally. They suggest another way of distinguishing criteria of improvement when they write that, "Pervading all of the theories [of psychotherapy] either explicitly or implicitly is the conception that the primary aim of therapy is improved adjustment of the individual, his level of adjustment being a function of his understanding of the self and the world, his handling of anxiety, and the effectiveness and social acceptability of his behavior" (6, p. 132). Miller (7) lists four criteria that have been used in the evaluation

of therapy: (a) the disappearance of symptoms, (b) the patient's social adjustment, (c) the patient's personal satisfaction, and (d) the therapist's satisfaction with the patient's adjustment. Miller considers these criteria as conflicting with one another, and therefore stresses the importance of studying to what extent and under what circumstances they are correlated.

Using the distinctions made by Kogan and Hunt and by Miller as a point of departure, the authors derived three criteria of improvement (10): 1. *Comfort*: The patient's stated purpose in coming to treatment is usually to increase his comfort in his daily functioning, and particularly to alleviate his symptoms. This conscious goal, of course, does not represent the complete picture of his motivations regarding therapy. Secondary gains derived from the illness, for example, may produce forces in the opposite direction. 2. *Effectiveness*: Those with whom the patient interacts in his daily life are usually most concerned with his effectiveness as a social being, his success in the performance of his social roles and in the management of his transactions with others. 3. *Self-awareness*: Most therapists, it can be assumed, are also concerned with increasing the patient's comfort and effectiveness. They tend to judge the effects of therapy, however, by reference to certain additional values, which are considered prerequisites to the attainment of greater comfort and effectiveness. These values include, for example, the acquisition of "insight," the correction of perceptual distortions, and the release of repressions. Self-awareness is one such value which is shared by a number of therapeutic approaches and was therefore selected as a third criterion of improvement.

It is reasonable to expect a certain amount of overlap between these three criteria, particularly since they may often involve evaluations of the same behavior from three different viewpoints. Nevertheless, it is quite likely that therapy will have differential effects on such

¹ The research for this paper was done at the Henry Phipps Psychiatric Clinic, Johns Hopkins Hospital as part of the Group Psychotherapy Research Project. Funds necessary to carry out the experimental work reported here were supplied by the Veterans Administration. The completion of the investigation was supported by a research grant from the National Institute of Mental Health. Herbert C. Kelman was a Public Health Service Research Fellow of the National Institute of Mental Health while participating in the study.

The writers are grateful to Jerome D. Frank, Director of the Group Psychotherapy Research Project, who made immeasurable contributions to this study and to the thinking behind it, and to Eduard Ascher, Stanley Imber, Earl Nash, David Rosenthal, and Anthony Stone, for their cooperation and suggestions in all stages of the research.

criteria. It cannot be simply assumed that a patient who improves in one respect necessarily improves in all other respects as well. Instead, it is necessary to establish the specific relationships between different kinds of changes and to study the conditions affecting these relationships. The criteria proposed here represent one way of classifying the different kinds of change that can be expected. The purpose of the present study was to investigate the interrelations among these criteria, using a small sample of patients participating in group therapy.

METHOD

Since the three criteria are very broad concepts, it is possible to devise a variety of measures for each, designed to tap different aspects of the criterion. It is conceivable that the kinds of relationships found differ with the specific operations defining any given criterion. In order, therefore, to represent some of the manifold aspects of each criterion, three measures of comfort, two measures of effectiveness, and two measures of self-awareness were used in the present study. It would be more accurate, consequently, to say that the purpose of the study was to investigate the interrelations among seven measures of improvement, representing (though not exhaustively) three separate criteria of improvement.

Subjects

The sample used in this study consisted of fifteen neurotic patients, seven men and eight women, who participated in one of three therapy groups at the Out-Patient department of the Henry Phipps Psychiatric Clinic for a period of at least 20 weeks. The therapeutic method was analytically oriented group therapy. Two of the groups were conducted by one psychiatrist and the third by another. Both therapists had had a minimum of 12 years of experience in therapy. Their orientation was eclectic, based largely on the theories of Sullivan, Freud, and Meyer. Each group met once a week for one hour. Further details about the sample, the form of therapy, and the therapists can be found elsewhere (9).

The patients were tested at the beginning of therapy and again after 20 weeks of treatment. The difference between the "before" score and the "after" score on each measure constitutes the index of change or improvement. A description of the seven measures follows.

Measures of Comfort

Self-satisfaction Q sort. Rogers (11) and others (e.g., 4) have used the degree of congruence between the patient's perceived self and ideal self, obtained by correlating two Q sorts, as an index of self-satisfaction. The same general method was used in this study, but special Q sort items, relevant to the group situation,

were devised.² Using Bion's concepts (1) as modified by Rosenthal (12) and Stock (13), six drives were distinguished which underlie interaction in group situations: pairing, counter-pairing, fight, flight, dependence, and counter-dependence. For each drive, ten items were selected describing behavior of a patient in a group judged to be characteristic of this particular drive (e.g., Fight: "I prolong or intensify arguments." Counter-dependence: "I resist suggestions of the doctor"). The patients sorted the 60 items twice: once to describe their characteristic behavior in the therapy group, and once to describe how they would ideally wish to behave. For each Q sort, sums of scores for each drive were obtained and rank ordered.³ The rank-order correlation between the Perceived Self Q sort and the Ideal Self Q sort of a given patient provided the measure of his satisfaction with his behavior in the group. The difference between the rho obtained after four weeks of therapy and the rho obtained after 20 weeks constitutes the measure of change in the patient's degree of self-satisfaction.

Symptom-disability checklist. A modification of the Cornell Index (16) was administered before therapy and again after 20 weeks. The list contained 41 physiological, behavioral, and affective symptoms. Patients were asked to check those symptoms which they experienced in the preceding month, and to rate on a three-point scale the degree to which each of these symptoms was disabling, i.e., interfered with their work, social activities, and over-all feeling of well-being. The total score was based on the number of symptoms checked, each weighted by its degree of disability. The difference between the "before" and "after" scores constitutes the measure of symptomatic relief. A patient does not improve in this measure simply by losing his presenting complaints, but by demonstrating a *net* decline in number and intensity of all symptoms.

Discomfort evaluation scale. This scale consisted of ten items, describing areas of the patient's behavior and especially of his interpersonal relationships in which he might experience discomfort, e.g., feelings of dependence, guilt, social uneasiness, etc. For each patient, these items were rated by each of three staff members who had seen him before therapy and again after 20 weeks: a psychiatrist conducting a psychiatric interview which concentrated on the patient's presenting symptoms, life history, and present situation; a social worker, conducting an interpersonal relationships interview (14); and a psychologist administering a Rorschach and TAT. The raters were asked to check each item which represented an area of discomfort for a given patient, and then to rate the degree of disturbance in this particular area on a three-point scale. The instructions to the raters spelled out a specific procedure designed to insure that each rating would be based on a

² The writers are greatly indebted to David Rosenthal for his extensive help in the construction of the Q sorts.

³ The test-retest reliability within a month for these rank orderings was computed for three randomly selected patients. The rho correlations for the Perceived Self Q sort were .93, .84, and .94 respectively; for the Ideal Self Q sort, .84, .84, and 1.00 respectively.

consideration and subjective weighting of all of the patient's significant interpersonal relationships on which information was available. After each staff member had completed his independent ratings, the three met together to discuss their ratings and to check interpretations derived from their separate methods against new evidence. This meeting yielded a common staff rating for each item, which was not a simple average of the three independent ratings, but often a new rating based on the pooled information from all three sources.⁴ The common staff ratings on all items were summed, thus yielding a total score on discomfort. Separate scores were obtained on the basis of pre-therapy tests and interviews and again on the basis of post-therapy tests and interviews. The differences between the "before" and "after" scores constitute the measure of change in degree of discomfort.⁵

Measures of Effectiveness

Sociometric ratings by group. It is commonly assumed that the therapy group gives the patient the opportunity to try out new responses and, in an environment relatively free of real life threats, to develop and test new social skills which he can transfer to his everyday relationships (3, 5). To the extent to which this is true, one would expect that the patient's increased effectiveness in the group would eventually be reflected in his interpersonal relationships outside the group. Accordingly, one of our measures of effectiveness was the average evaluation of each patient by his fellow group members on a modified sociometric questionnaire. Each patient was asked to rate all other patients in his therapy group on three dimensions: the extent to which he respected them, regarded them as leaders, and desired to be friends with them. The ratings were made on a four-point scale, and were assigned scores from zero to three. To arrive at a measure of effectiveness for a given patient, the average rating he received on each dimension was computed and the three averages were then summed. The difference between the score

thus obtained after five weeks of therapy and the score obtained after 20 weeks constitutes the measure of change in effectiveness. It must be stressed that this measure is based on changes in effectiveness *in the therapy group* rather than in the community. Increased effectiveness in the group is not necessarily accompanied by corresponding changes in daily life. Conversely, for some patients, behavior in the community may become more effective and yet behavior in the group may remain unchanged or even deteriorate. To obtain more direct measures of effectiveness in the community would, of course, require different sources of judgment and possibly different instruments (10).

Ineffective behavior evaluation scale. This scale consisted of sixteen items, describing kinds of difficulties a patient might experience in achieving goals that are important to him and in carrying out the social roles which he is expected to assume, e.g., difficulties due to submissiveness, indecisiveness, inability to control feelings, inadequate use of capacities, etc. The three staff members who filled out the discomfort evaluation scale also completed this form, checking each item which represented an area of difficulty for a given patient and then rating on a three-point scale the degree to which this difficulty is characteristic of the patient. The procedures for making the ratings, for arriving at common staff ratings, and for developing a total score, were the same as those used for the discomfort evaluation scale.

Measures of Self-awareness

Self-awareness Q sort. Self-awareness was defined as the extent to which a patient sees himself as others see him and was measured by the degree of congruence between the patient's perceived self and his behavior as described by trained observers. To determine a patient's degree of self-awareness, the Q sort items described earlier were used. The Perceived Self Q sort was the same as the one used for measuring self-satisfaction. The Observed Behavior Q sort was obtained by asking observers to sort the 60 items for each patient in such a way as to describe his behavior in the group.⁶ The observers were two clinical psychologists and one social worker, each of whom was assigned to one of the three groups. Each had considerable training in group observation. Moreover, they were completely outside the therapy group and played no role in the therapeutic process. For these reasons, it was assumed that their perceptions of the patients' behavior in the group were relatively "objective." For each sorting, the six drives were rank ordered and the rho correlations between the Perceived Self and the Observed Behavior Q sorts were computed. The difference between the rho obtained after four weeks of therapy and the rho obtained after 20 weeks constitutes the measure of change in degree of self-awareness. A serious limitation on this measure is the fact that the observers had seen the patients outside the therapy group in order either to test or interview them; also, they observed the groups throughout

⁴ Since it was suspected that the psychiatrist might be exercising a greater influence on the final staff rating than the psychologist or social worker, a comparison was made between the final staff ratings and each of the three independent ratings. This analysis showed that there was no overall tendency to adopt the ratings of any one of the three. It can be concluded, therefore, that the staff ratings do in fact represent a joint product.

⁵ It was not possible for the staff members to fill out the pre-therapy evaluation scale and to hold the joint staff sessions until after the post-therapy tests and interviews had been administered. Although the ratings were based on the records of the pre-therapy tests and interviews, it is quite likely that they were affected by the post-therapy tests and interviews. The before- and after-scores derived from the evaluation scale are, therefore, not completely independent. A cautious interpretation of the change scores would be that they are essentially *ratings of change* rather than measures of change based on the patients' status before and after therapy. The same limitation holds true for the ineffectiveness evaluation scale described here.

⁶ The test-retest reliability was obtained by asking each observer to repeat the sorting on a randomly selected patient of his group. In each case, the rho correlation was 1.00.

TABLE 1
COMPARISON BETWEEN MEAN BEFORE-SCORES AND MEAN AFTER-SCORES ON DIFFERENT MEASURES OF IMPROVEMENT ($N = 15$)

Measure	Lowest Score Possible*	Highest Score Possible*	Mean Before-score	Mean After-score	p-Value for Difference**
Measures of comfort					
(1) Self-satisfaction Q sort	-1.00	1.00	.44	.41	—
(2) Symptom-disability checklist	123	0	44.13	36.33	<.01
(3) Discomfort evaluation scale	30	0	22.14	17.20	<.01
Measures of effectiveness					
(1) Sociometric ratings by group	0	9	5.47	5.21	—
(2) Ineffective behavior evaluation scale	48	0	16.27	13.60	<.01
Measures of self-awareness					
(1) Self-awareness Q sort	-1.00	1.00	.43	.44	—
(2) Sociometric ratings predicted for self	9	0	2.88	2.63	—

* By "lowest score possible" is meant the extreme of the possible range of scores on the unfavorable end, i.e., the end denoting "poor mental health." Movement away from the lowest score possible in the direction of the highest score possible would be considered "improvement."

** The p values were obtained by using Wilcoxon's (15) nonparametric method for determining the significance of the difference between paired replicates.

the 20 weeks, so their final sortings were inevitably based on a larger sample of behavior than the initial sortings.

Sociometric ratings predicted for self. In filling out the sociometric questionnaire described above, patients were also asked to predict the ratings they would receive on each dimension from each of their fellow group members. For each dimension, the average discrepancy between the ratings the patient expected and the ratings he actually received was computed. The three averages were summed and the score thus obtained was used as a measure of another aspect of self-awareness: the patient's awareness of the kind of impression that he makes on others. The difference between the discrepancy score obtained after five weeks of therapy and the score obtained after 20 weeks constitutes the measure of change in accuracy of prediction.⁷

The seven measures can be grouped into "indirect" and "direct" measures of improvement. Those derived from the evaluation scales are indirect in that they are based on staff ratings. Those derived from the Q sort, the sociometric, and the symptom checklist are direct in that they are based essentially on the patient's "performance" in relevant test situations. Of the direct measures, those derived from the Q sort and the sociometric have as their referent the therapy group; they measure the patient's satisfaction with, effectiveness in, and awareness of, his group behavior. One cannot assume, of course, that behavior in the therapy group is identical with behavior in the community, especially during the

early stages of therapy. Performance in the group, however, can serve as a reasonable index of changes in the patient, particularly since the focus of interest here is to study interrelations among different measures rather than to evaluate a given form of therapy.

Three additional measures, which might help in providing dynamic explanations of patient behavior, were obtained. Each of the staff independently rated each of the patients with regard to (a) the degree of increase or decrease of anxiety displayed by the patient in the course of therapy, (b) the degree to which the patient tended to accept or reject repressive solutions to his problems, and (c) the degree to which the patient was self-revealing and discussed personal material.

RESULTS

Table 1 summarizes the data on the amount of change that took place between the before-test and the after-test on each measure. It should be noted that these data do not constitute a proper basis for evaluating group therapy since no control group was used. They are presented here in order to provide a basis for interpreting the interrelations among changes found on the different measures.

Three of the seven measures show significant change for the group as a whole: the symptom-disability checklist, the discomfort evaluation scale, and the ineffective behavior evaluation scale. It is interesting to note that these are the three measures which are based on some kind of rating of well-being by either the patient or the staff. The lack of significant

⁷ A third measure of self-awareness, a distortion evaluation scale (containing four items), was also used. The range of change scores which this measure yielded was extremely small, however; 0, 1, or 2 were the only scores represented. Since it would be almost impossible to evaluate and interpret correlations based on such a distribution, it was decided to omit this measure.

TABLE 2
INTERRELATIONS AMONG CHANGE SCORES ON SEVEN MEASURES OF IMPROVEMENT ($N = 15$)^{*}

	Self-satisfaction Q Sort	Symptom-disability Checklist	Discomfort Evaluation Scale	Sociometric Ra- tings by Group	Ineffective Be- havior Evalua- tion scale	Self- awareness Q Sort
Symptom-disability check- list	.14					
Discomfort evaluation scale	.16	.44				
Sociometric ratings by group	-.36	-.10	.07			
Ineffective behavior evalu- ation scale	.02	.36	.44	-.08		
Self-awareness Q sort	.11	.54**	.18	-.31	.44	
Sociometric ratings pre- dicted for self	.10	.14	.03	-.29	-.30	.02

^{*} In the computation of Pearson r 's, the scores on each measure were so arranged that all would move in the same direction, i.e., the higher the score, the greater the improvement. A positive correlation would mean, therefore, that individuals who improved according to one measure also tended to improve according to the other.

^{**} $p < .05$.

over-all changes on the other measures does not necessarily mean that no changes occurred. A review of the data shows that some patients changed in a positive direction and others in a negative direction, and that these two tendencies tended to cancel each other out. An example of this type of variability is found in an analysis of patients who showed an increase and those who showed a decrease in self-awareness. When the self-awareness Q sort data were compared with the staff ratings on changes in patient anxiety level, it was found that patients whose anxiety decreased in the course of therapy became more aware of their own behavior, whereas those whose anxiety increased became less aware.

Table 2 presents the correlations between change scores on all pairs of measures. It can be seen that of the 21 correlation coefficients listed in the table, only one reaches significance at the .05 level, the r between the self-awareness Q sort and the symptom checklist. There are six negative (though nonsignificant) correlations, five of them involving the sociometric ratings by the group.

This over-all picture indicates that, at least as far as the present measures are concerned, there is no support for the assumption that changes on different measures go together and that improvement can be considered a unitary process. Even the four measures that refer specifically to the patient's behavior in the therapy group—those based on the Q sort and the sociometric—reflect this independence among criteria despite the fact that these four measures are all based on the same limited

range of group behavior which maximizes the likelihood of significant correlations. In short, these data underscore the necessity for specifying the criteria and measures used in evaluation whenever results of psychotherapy are reported.

The single significant correlation that was obtained obviously has to be viewed with caution, since it could easily be a chance phenomenon. It is tempting to speculate about its meaning, however. One hypothesis that might explain the correlation is that both the symptom checklist and the self-awareness Q sort are responsive to reduction in anxiety. To the extent to which the patient's anxiety decreases, his symptoms will tend to disappear. Also, his self-awareness will tend to increase, since there will be less of a need for a defensive self-concept. Some suggestive evidence for this hypothesis was obtained from the staff ratings of changes in patient anxiety: those patients who were rated as "decreased in anxiety" showed the greatest change both on the symptom checklist and on the self-awareness Q sort.

DISCUSSION

The findings of the present study seem to provide no evidence for the assumption that changes on different measures go together. The over-all picture presented in Table 2 makes it impossible to reject the hypothesis of "no relationship." The status of this finding, however, must remain ambiguous because of questions of reliability and of inter-

pretation which cannot be answered by the present data.

First, there is some ambiguity regarding reliability. There are at least two statistical reasons why the finding of "no relationship" might fail to show up in a replication of this study. It may be that positive though small relationships exist, but that a larger sample is necessary in order to demonstrate their significance. Furthermore, on four of the seven measures there was no significant change for the group as a whole over the period of therapy (see Table 1). While it is possible, of course, that some of the patients showed real changes even if the group as a whole did not, it is also possible that the obtained changes merely represent measurement error. If the latter is true, then one can conclude very little from a lack of relationship between measures. There is, after all, no reason to expect a correlation between two sets of random errors. It can be argued that if real changes had occurred, they might very well have been related to each other. Where there was no significant change to begin with, measurement error cannot be rejected as a possible explanation.

In view of these limitations, one can only draw the modest conclusion that the data do not confirm and hence throw doubt on the frequent assumption that improvement is a unitary process.

Secondly, even if there were no question about the reliability of the finding of "no relationship," the interpretation of this finding would still be ambiguous. An over-all lack of relationship among measures would be found if the measures were completely independent. It would also be found, however, if the measures were related to each other in a complex way. That is, if under some conditions (or for certain types of patients) two measures are positively correlated and under other conditions they are negatively correlated, then an over-all correlation of around zero is likely to result.

The second interpretation provides a model for analysis and hypothesis building which is probably more meaningful from a clinical point of view. It is hard to believe that changes in self-awareness, for example, will be completely unrelated to changes in effectiveness. On the other hand, it is quite reasonable to expect that for some patients increased self-awareness will go with increased effectiveness,

while for other patients it will go with reduced effectiveness. Which interpretation is correct has to remain ambiguous for the moment. Either interpretation, however, points to the need for paying attention to the specific criteria and measures that are used in evaluation. At the very least, it seems likely that not *all* measures of improvement go together or—if the second interpretation is correct—that not all measures go together for all patients and under all conditions. Further research would have to indicate what the relevant conditions are. Such knowledge would help not only in the selection of appropriate therapeutic programs for particular kinds of patients, but also in the understanding of the therapeutic process.

The present study has suggested some hypotheses about the conditions under which positive correlations between changes on two measures would be expected, and the conditions under which negative correlations would be expected. In general, it would appear that the kinds of changes that occur in group therapy can be related to the activities and experiences of the patient in the group. The correlations between changes, therefore, can also be understood in terms of these activities and experiences.

Consider, for example, the relationship between changes in symptoms and changes in self-satisfaction. One may hypothesize that for "under-controllers," i.e., patients whose ego-control is overly flexible, who tend toward variable and impulsive behavior, and who are quite conscious of their emotional difficulties and their low self-satisfaction (2), the correlation would be positive. As these patients gain greater control over their impulses and learn to stabilize their behavior, they will become more satisfied with themselves and also lose their presenting symptoms. For the "over-controllers," on the other hand, i.e., patients whose ego-controls are rigid, who deal with their difficulties in a defensive and repressive manner, and who deny that there is a discrepancy between their perceived self and ideal self, a negative correlation may be hypothesized. These patients will tend to lose their presenting symptoms only as they learn to relax their controls and to adapt to changing conditions. To the extent to which they relax their controls, however, their defensive self-satisfaction will tend to diminish. Thus,

as the patient improves symptomatically, his self-satisfaction may go down. Suggestive evidence for this hypothesis was obtained when staff members rated each patient in terms of his tendency to be "self-revealing" (typical of the "under-controller") as against a tendency to "hold back" (typical of the "over-controller"). For the patients who were rated as self-revealing, there was a significantly higher correlation between the two measures than for the patients who held back.

Aside from personality factors, such as ego-control, the nature of the patient's group experience is likely to affect changes and hence correlations between changes. The correlation between changes in self-awareness and changes in self-satisfaction, for example, may be positive if the patient is given warmth and support in the therapy group. For such a patient "to become more objective about his own behavior" may mean to learn (perhaps for the first time in his life) that others really like him. As his self-awareness increases, therefore, he would also become more satisfied with his interpersonal behavior. On the other hand, if the patient is criticized and rejected by other members of the group, one would expect a negative correlation between the two measures. For such a patient, "to become more objective" may mean to learn how unacceptable to others he really is. Thus, as his self-awareness increases, he may become more dissatisfied with himself.

The above example points to the possible negative or "anti-therapeutic" effects that group therapy may have. These possibilities become particularly apparent when the patient's effectiveness in the group (defined by sociometric ratings in the present study) is used as a measure of improvement. A great deal depends on who the other group members are and what kinds of group standards have developed. For example, in some groups a patient who has insight into his own behavior may alienate other patients by threatening their defenses.⁸ Here one would expect a negative correlation between changes in self-awareness and changes in effectiveness in the group. In those groups, however, in which

insight is encouraged, a positive correlation would be anticipated.

To take another example, one might hypothesize that increases in self-satisfaction depend on the extent to which the patient engages in frank interaction with other patients and expresses his true feelings. In therapy groups in which this kind of free expression is generally accepted, it would also be expected to lead to increased effectiveness in the group. In groups in which such expression is not sanctioned, however, it may antagonize the other patients and hence lead to reduced effectiveness in the group.

Finally, the group norms may be such as to accelerate the drive towards improvement. A patient who improves may receive the support and encouragement of his fellow patients, who in their turn are encouraged by his progress. Where that is true, a positive correlation would be expected between changes in symptoms and in group effectiveness. It is also possible, however, that patients mutually reinforce their defenses and develop group norms favoring a "slow-down" in improvement.⁹ The more disturbed a patient is, the higher may be his status in the group. The patient who loses his symptoms may also lose his status; he may constitute a threat to the others and therefore be rejected by them.

As can be seen from Table 2, there is a slight tendency for the change scores on sociometric effectiveness to be negatively correlated with the other measures. This would suggest that some of the antitherapeutic agents discussed may have been present in the groups studied, at least during the early phases of therapy. It is important to keep these possibilities in mind; one cannot assume that patients in a group will always be accepting and supporting of one another. Such negative possibilities exist in all forms of therapy, and it is important to identify them in order to make possible a rational choice between different forms of therapy for a given patient.

SUMMARY

The purpose of the present study was to investigate the interrelations among seven meas-

⁸ This possibility is suggested in a study by Norman (8), using a non-patient population. He found that less accepted than those with a medium degree of self-insight.

⁹ In this respect, it is interesting to compare the therapy group with the industrial work group. Numerous studies have shown that group standards can have the effect of increasing or decreasing an individual's output.

ures of improvement in therapy, representing three separate criteria of improvement: comfort, effectiveness, and self-awareness. Comfort was measured by a self-satisfaction Q sort, a symptom-disability checklist, and a discomfort evaluation scale. Effectiveness was measured by sociometric ratings and an ineffective behavior evaluation scale. Self-awareness was measured by a Q sort and predictions of sociometric ratings. Scores on these seven measures were obtained for fifteen neurotic patients at the beginning of a course of group therapy and again after 20 weeks of therapy. The change between before- and after-scores in each case constituted the measure of improvement.

Significant change was found on three measures: the symptom checklist and the two evaluation scales. Of the 21 intercorrelations between change scores, only one was significant at the .05 level. It was concluded that, at least as far as the present measures are concerned, there is no support for the frequent assumption that changes on different measures and criteria go together and that improvement can be considered a unitary process. The need for specifying the particular criterion and measure used for evaluation of psychotherapy was stressed.

Some of the ambiguities regarding reliability and interpretation of the findings were discussed. It was suggested that some of the near-zero correlations found may be due to the fact that under some conditions (or for certain types of patients) two measures may be positively correlated, while under other conditions these same measures may be negatively correlated. Some hypotheses about the conditions that may affect the direction of correlations were offered. They included hypotheses about the role of personality factors, such as ego-control, and the role of group experiences, such as the amount of support or rejection a patient receives from others and the kind of group standards that evolve.

REFERENCES

1. BION, W. R. Experiences in groups. *Hum. Relat.*, 1948, **1**, 314-320, 487-496; 1949, **2**, 13-32, 295-304; 1950, **3**, 3-14.
2. BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254-259.
3. FRANK, J. D. Group methods in psychotherapy. *J. soc. Issues*, 1952, **8** (2), 35-44.
4. HANLON, T. E., HOFSTAETTER, P. R., & O'CONNOR, J. P. Congruence of self and ideal self in relation to personality adjustment. *J. consult. Psychol.*, 1954, **18**, 215-218.
5. KELMAN, H. C. Two phases of behavior change. *J. soc. Issues*, 1952, **8** (2), 81-88.
6. KOGAN, L. S., & HUNT, J. McV. After comment. *Psychol. Serv. Center J.*, 1950, **2**, 132-138.
7. MILLER, J. G. Objective methods of evaluating process and outcome in psychotherapy. *Amer. J. Psychiat.*, 1951, **108**, 258-263.
8. NORMAN, R. D. The interrelationships among acceptance-rejection, self-other identity, insight into self, and realistic perception of others. *J. soc. Psychol.*, 1953, **37**, 205-235.
9. PARLOFF, M. B. Some factors affecting the quality of therapeutic relationships. *J. abnorm. soc. Psychol.*, 1956, **52**, 5-10.
10. PARLOFF, M. B., KELMAN, H. C., & FRANK, J. D. Comfort, effectiveness, and self-awareness as criteria of improvement in psychotherapy. *Amer. J. Psychiat.*, 1954, **111**, 343-351.
11. ROGERS, C. R. Studies in client-centered therapy: III. The case of Mrs. Oak—a research analysis. *Psychol. Serv. Center J.*, 1951, **3**, 47-165.
12. ROSENTHAL, D. *Perception of some personality characteristics in members of a small group*. Unpublished doctor's dissertation, Univer. of Chicago, 1952.
13. STOCK, DOROTHY. *The relation between the sociometric structure of the group and certain personality characteristics of the individual*. Unpublished doctor's dissertation, Univer. of Chicago, 1952.
14. VARON, EDITH, & MAAS, H. S. Evaluating social adjustment through a systematic study of interpersonal relationships. *J. psychiat. soc. Wk.*, 1951, **21**, 3-12.
15. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Company, 1949.
16. *Cornell Medical Index Health Questionnaire*. New York: Cornell Univer. Medical College, 1949.

Received July 19, 1956.

DETERMINANTS OF CONFORMITY AND ATTITUDE CHANGE¹

KENNETH R. HARDY

Brigham Young University

An adequate understanding of human social behavior requires the simultaneous study of the nature of social influence and the causes of individual variations in susceptibility to that influence. A notable example of the recent trend to study these concomitantly is seen in the work of Asch on independence and submission to group pressure (2, Ch. 16). The study reported here had two objectives: First, to extend the work of Asch from perceptual to attitudinal judgments, from public conformity to attitude change as well. Specifically, this objective was to compare the effects upon conformity and attitude change of unanimous group pressure as contrasted with near-unanimous group pressure. The second objective was to investigate the role of affiliative motivation in affecting susceptibility to group normative pressure.

Attitude change is here considered to include any change in the organization of "motivational, emotional, perceptual, and cognitive processes with respect to some aspect of the individual's world" (10, p. 152). Attitude change arises from changes in perceived social reality, or from other influences which are thought to reflect "objective" reality dependably, such as the opinions of experts or other prestigious persons, the evidence of research, personal experience, and the like. The importance of "social reality" (6, 16), defined in terms of the shared consensus of (significant) people in the person's environment, is now widely acknowledged as a determinant of social attitudes.

Conformity is here defined as the public avowal of a belief or attitude at variance with

one's prior position, which avowal tends to correspond to the position approved by the group in which the avowal occurs. It may occur as an accompaniment of attitude change based upon a shift in perceived social reality. It may also occur in the absence of attitude change in situations where it is believed to be instrumental to the satisfaction of such motives as sex, achievement, affiliation, power, or hunger.

The type of social situation under investigation is one in which a small group of men meet informally. They state their opinions on a given topic, expressing unity of opinion. What will a person do whose position on the issue has hitherto been at variance with that of the others? Will he change his mind about the issue, or will he, at the least, publicly express agreement with the group? What will he do if another person in the group expresses a similarly deviant position? The predictions investigated were as follows: First, there would be (a) greater conformity and (b) greater attitude change under conditions of unanimous opposition to the subject's prior position, than where the subject has a "partner" who shares his opinion. This is presumed because in the "partner" situation the pressures toward conformity are less, and the criterion of social reality loses much of its force when there is a difference of opinion expressed within the group. Secondly, there would be (a) greater conformity and (b) greater attitude change by those who have greater affiliative motivation than by those with lesser affiliative motivation. This prediction is based partly on the greater desire of the former to be accepted and well thought of, combined with the common realization of the tendency of groups to reject those deviant from group norms (17), and partly on their presumably greater intellectual dependence upon social reality.

METHOD

The fundamental plan of the research was to submit a subject to a social-influence situation wherein his own attitude would be opposed by all, or alternatively, all but one of the other persons present. The focus of study was the S's public response to this situation—and his private (attitudinal) response.

¹ This report is based upon a doctoral dissertation, University of Michigan, 1954, which is on file with University Microfilms, Ann Arbor, Michigan, as publication No. 8311 (price \$1.56), and which contains the raw data, instruments used, and a fuller discussion of theory, method, and results.

The author wishes to express his appreciation to Dr. Daniel Katz and the other members of his committee for their counsel and encouragement. He is also grateful to Dr. J. W. Atkinson, Dr. J. D. Birch, Joseph Veroff, and Robert Hamblin for assistance in special phases of the research.

In order to accomplish this objective, a three-stage design—of pretest, experimental manipulation, posttest—was used. Male Ss were recruited from classes in introductory psychology and general chemistry at the University of Michigan. They were asked to volunteer two hours of their time to a study of the relative efficacy of various methods of measuring attitudes. The attitude to be studied was that toward divorce, and considerations of control and expedience led to the elimination of all Catholic, married, and Negro volunteers.

Pretest. The pretest, administered to small groups of Ss at a time, consisted of four instruments (discussed below), all ostensibly representing methods of measuring attitudes. Actually, the major purpose of the pretest was to secure data on the Ss' affiliative motivation and attitudinal position respecting divorce so that they could be assigned to the appropriate experimental variation.

The first instrument given was the TAT-type measure of affiliative motivation (n Affil) developed by Atkinson, Heyns, Shipley, and Veroff (3, 19). The standard story blanks were distributed and standard instructions for administering a projective measure of motivation were then read (cf. 13). Four pictures were shown in the following sequence: (a) two men conversing; (b) group of seven men around a table; (c) older woman seated, looking away from young man by window; (d) man by lamppost (TAT 20).

It was intended to structure a situation that would be only mildly arousing of affiliative motivation, so that affiliation imagery would be aroused mostly by the pictures shown. The small groups tested consisted of strangers or relatively casual acquaintances meeting in a seminar room on campus. The testing was usually preceded by a minor amount of "small talk" as the men arrived in the room. This situation thus appeared to be about as affiliatively arousing as the "neutral" condition of Shipley and Veroff (19), a conjecture that is substantiated by a comparison of scores obtained in the present study with those obtained on a similar population by these investigators.

Following the completion of the stories, Ss were asked to write their names on the back of each sheet. Assurances were given of the confidential treatment of this and other material collected during the study. The blanks were then collected, and a questionnaire booklet containing the other three instruments was distributed.

The first questionnaire was the California F scale (1, pp. 255-257), introduced as dealing with general public opinion. This scale, and the self-ratings discussed below, were primarily used to bolster the facade that the study was investigating various methods of measuring attitudes, at the same time drawing attention away from the divorce issue. They were also used to explore possible relationships they might have with the experimental variables.²

The second questionnaire, which pertained to the divorce issue, was introduced as dealing with a specific

attitude quite intensively rather than dealing superficially but extensively with many attitudes as the preceding questionnaire had done. The first page asked the S to write "your attitude and feelings toward divorce and the reasons you feel the way you do about it." He was allowed about ten minutes to do this. The objective of this section was to provide arguments that his supporter could use in the group situation. The next section consisted of three parts. The first part was a nine-point graphic scale of leniency toward divorce, on which S indicated his own position and those of his parents and friends. This scale was also used in the group-influence situation, where S had to state his position publicly. A comparison of the two positions taken by S provided the operational definition of conformity. The second part consisted of 12 attitude statements dealing with various aspects of divorce: the nature of the marital contract, divorce as a solution to an unhappy marriage, effect of divorce upon children and society generally, the degree of present abuse of divorce, and the obligation partners should feel to remain married. There were six "liberal" and six coordinate "conservative" statements. This set of statements, also given after the group situation, provided the operational definition of attitude change. For example, if an S marked "-" (mildly disagree) to a given statement initially, but afterward marked "+" (mildly agree), this would be scored as a change of two units. If he marked "++" (strongly agree) initially and "-" later, three units of change would be recorded. If the change occurred in the direction of the group influence, it was scored positively, and if in the opposite direction, scored negatively. The algebraic sum of these shifts over the twelve statements was the S's attitude-change score. The third part of this questionnaire contained two questions that assessed the strength and certainty of the S's attitude toward divorce.

The final instrument in the questionnaire booklet was a series of ratings for 23 personality traits. Most of the traits were adapted from Fiske (7), who in turn adapted them from a longer list used by Cattell (4). In Fiske's factor-analytic study, seven of the self-rated traits had loadings of .40 or better on a centroid factor he labeled "conformity."

Experimental manipulation. The group-influence situation came after an interval of four to eleven weeks following the pretest. In the interim, the n-Affil scores were computed and the resulting distribution divided into high and low n-Affil groups.³ These groups were then assigned to the *partner* or *no-support* variations so as to control for initial position on the divorce issue (as indicated by the graphic scale) and time elapsing since the pretest. Each S met with a group of six other students, all of whom were presumably Ss. Actually these six were confederates who had been instructed to give standardized "arguments" using five kinds of appeals: emotional, personal contact or experience, scientific evidence, appeal to authority, and acceptability of change as a result of further thought and

² Analysis of results showed no relationships between the F scale and either conformity or attitude change. Negative results were also found between the "conformity" factor of the self-ratings and these dependent variables.

³ Reliability was established by comparing the author's scoring of two samples of stories obtained by previous research with the pooled scoring of the authors of the n-Affil manual (7) on these same stories. The average rho for the two samples was .80 for total scores.

investigation. The content of each argument centered around some aspect of divorce sampled by the attitude statements in the questionnaire.

As an example, the antidivorce argument appealing to authority and dealing with the effectiveness of divorce as a solution to an unhappy marriage is as follows: "Some people say that a couple ought to get divorced if they are incompatible. A psychiatrist who is a friend of the family told me that the real reason marriages are unhappy is that the couples have not really learned how to solve their problems. The definition of a happy marriage is not one that has no differences or problems; it is one where the husband and wife know how to solve their problems as they arise. Getting divorced and later remarrying doesn't help people to learn these techniques. What they need to be happy is to learn how to get along together." The confederate was free to state the argument in his own words.

It should be noted that this technique of not only stating a judgment but also buttressing it with "reasons" represents a methodological innovation, one that represents the ordinary social-influence situation more faithfully than the perceptual situations studied by Asch (2) and Sherif (18), in which the judgment is immediately given and there is no room for discussion.

In the group situation, *Ss* were seated in a semi-circle (the confederates were free to take any seat) facing the *E*, who began by stating,

During the first hour we worked with several methods of measuring attitudes. Today we are going to work with another method which I might call a "group interview." What I shall do is call on each of you in turn and ask you to tell me how you feel about a particular topic and why you feel the way you do about it. The topic is divorce. I shall write down in summary form what you say, and then attempt to restate it to see if I have it accurately.

After each interview was completed, *E* held up a large placard on which was the graphic scale used during the pretest. After the person stated his position, together with any reasons for placing himself there, *E* called on the next individual in his list of names. This procedure was repeated until all the confederates, and finally the *S*, had spoken.

In the no-support variation, all of the confederates opposed the *S*'s previously indicated position. Since two sets of arguments had been constructed, one favorable, the other unfavorable to divorce, the set used was the one opposed to *S*'s position as indicated on the pretest. A position on the graphic scale paralleled each argument. The arguments were always given in the same sequence. In each case the fifth man "agreed" with the others but added no arguments to those already given. In the partner variation, the fifth man opposed the others, giving arguments the *S* had used during the pretest; he placed himself at the *S*'s initial position on the graphic scale. By this procedure the amount of new ideas presented to the *S* was kept constant across the variations. The fifth man was used so that the group norm would be clearly established by the first four men and maintained by the sixth person.

Upon completion of the "group interview," *E* said, "Now there is one other thing I'd like you men to do before we are through and that is to fill out a short

questionnaire for me individually. In order to do this, I will assign you to separate rooms." Each man was handed a (posttest) questionnaire as he went to his room, but the *S*, of course, was the only one who filled it out.

Posttest. This questionnaire, which was filled out privately, contained measures of liking of the group, perception of the group situation, and repeat measures of the attitude toward divorce. Upon completion of the questionnaire, *E* interviewed the *S* and explained to him the real nature of the study. At this time *S* was given the opportunity to serve as a confederate (which most *Ss* were eager to do, especially those most affected by the group situation). The major source of confederates, however, was volunteers from classes in social psychology and human relations.

Although a total of 45 "group interviews" were conducted, in 13 cases the experimental conditions were not met by the failure of one or two confederates to appear, and in one case there were two supporters due to a mix-up in instructions. The results are based upon the 31 remaining cases.

RESULTS

The first question to be answered is, "Did the situation really provoke conformity and changes in attitude?" With regard to conformity, 14 men (45%) shifted two or more scale points on the graphic scale, in the direction of the group influence. (The average distance from initial position to the group norm was 4.36 scale points.) In contrast, only three men shifted in the opposite direction, and none of them changed more than one point. Similarly, 14 men had a positive attitude-change score of six or more points; that is, they changed their position on at least half of the attitude statements (or the equivalent). Only two men had a negative score this high. It should be noted here that the correlation between conformity and attitude change was .51. Observations of these men confirm those of Asch in detecting more or less severe signs of emotional activation in the group situation. One *S*, for example, spoke very rapidly (atypical of him) in pronouncing his agreement with the others, then blurted, "Boy, it's hot in here; can't we open a window?" He was the only one to notice the "heat"!

Effect of social support. Results of *t* tests indicated that the absence of social support was significantly related to attitude change, but the relationship to conformity fell short of significance, as shown in Table 1. The hypothesis regarding the effect of social support upon attitude change thus received confirma-

TABLE 1

EFFECT OF SOCIAL SUPPORT ON CONFORMITY AND ATTITUDE CHANGE

Dependent Variable	Condition			
	No support (N = 16)	Partner (N = 15)	t	p
Conformity				
Mean	1.81	1.16	1.30	.11
SD	1.55	1.12		
Attitude change				
Mean	6.97	2.00	1.79*	<.05
SD	9.44	5.05		

* A formula recommended by Dixon and Massey (5, pp. 104-105) was used here because of heterogeneity of variance between the two groups.

TABLE 2

EFFECT OF n AFFILIATION AND SUPPORT ON CONFORMITY

a. Conformity in Variations of n Affiliation and Support

Degree of support	n Affiliation		
	High	Medium	Low
No support			
Mean	1.97	2.97	-.175
SD	1.41	.58	.43
N	6	6	4
Partner			
Mean	.06	1.83	1.58
SD	.76	1.15	.59
N	5	5	5

b. Results of Variance Analysis

Source	df	Mean Square	F
Support	1	1.51	1.49
n Affiliation	2	9.88	9.77
Interaction	2	20.85	20.62
Within cells	25	1.01	

 $F_{.01} = 5.57$ for 2 and 25 df

tion, but the evidence is ambiguous respecting the effect of support upon conformity.

Effect of affiliative motivation. A comparison of high and low n-Affil groups showed no differences between these groups on either conformity or attitude change. However, it seemed possible that social support might be interacting with n Affil in such a way as to cancel out differences between the affiliation groups, or that an alternative division of the n-Affil distribution might show positive relationships.

This second possibility recommended itself from some research done on n Achievement

TABLE 3

EFFECT OF n AFFILIATION AND SUPPORT ON ATTITUDE CHANGE

a. Attitude Change in Variations of n Affiliation and Support

Degree of support	n Affiliation		
	High	Medium	Low
No support			
Mean	6.75	11.00	1.25
SD	7.37	11.37	3.36
N	6	6	4
Partner			
Mean	.40	1.20	4.40
SD	4.84	4.26	5.08
N	5	5	5

b. Results of Variance Analysis

Source	df	Mean Square	F
Support	1	153.78	2.51
n Affiliation	2	43.44	
Interaction	2	242.85	3.96
Within cells	25	61.35	

 $F_{.05} = 3.38$ for 2 and 25 df

by Lowell (12), McClelland and Liberman (14), and Moulton (15), who found more clear-cut relationships with certain perceptual and behavioral phenomena on the basis of a high-medium-low division than with a dichotomization of the n-Ach distribution. They found the high group, which is conceived to be positively motivated toward achievement, to suffer no decrement in performance nor perceptual defense under achievement-arousing conditions. The middle n-Ach group, which is conceived to be ambivalently motivated toward achievement, *does* show defensiveness to perception of failure words, and a performance decrement under arousal conditions.

Interaction of social support and n Affil. To examine these two possibilities, analyses of variance were made, using the two categories of social support, and three categories of n Affil based upon a division of that distribution into three groups as nearly equal in size as possible. Table 2a reports the *N*s, means, and *SD*s of the six resulting cells for conformity; Table 3a the corresponding data for attitude change. Before analyses of variance were made, the cell frequency (with the accompanying sums of squares) of the "Low-n-Affil, No-support" cell was adjusted to make possible an analysis procedure based

upon proportional cell frequencies (cf. 11, pp. 151-155). Also, before each analysis, the assumption of homogeneity of variance within cells was checked by using Bartlett's test, and was verified in each case. Table 2b gives the results of the variance analysis for conformity, Table 3b for attitude change.

The results for conformity indicate that both *n* Affil and the interaction component are highly significant factors, but support is not. Within the table the following results were obtained:

1. The effect of *n* Affil is significant within each support variation. In the no-support condition, the high and medium groups are significantly more conforming than the low group. In the partner variation, however, the medium and low groups are significantly more conforming than the high group.

2. If the support variations are combined, then the medium *n*-Affil group is more conforming than either the high or low groups.

3. The no-support variation is more conformity-producing for the high-need group and less conformity-producing for the low-need group than the partner variation. There is no difference between the support variations for the middle-need group.

From these results it would appear that conformity is a joint function of the social conditions and the affiliative motivation of the individual. For those high in *n* Affil the presence of a partner results in the disappearance of conformity that is definitely present under conditions of no support. Those in the middle range of the *n*-Affil distribution do not drop significantly in conformity with the presence of a partner. Those in the low *n*-Affil group, who do not conform at all under conditions of no support, show a significant *increase* in conformity when a partner is present.

In regard to the other dependent variable, attitude change, somewhat similar fluctuations in mean values occur, but none of the factors is statistically significant. (If the between-groups variance is tested before the interaction component is accepted as significant (cf. 11, p. 166), the resulting *F* ratio is 2.37, which leaves the interaction of doubtful significance.)

Though there are somewhat similar patterns for conformity and attitude change, there is one important difference. The medium-need group conformed whether or not a partner was present; but they tended to change less in atti-

tude in the partner situation than in the no-support situation (*p* about .05).

DISCUSSION

These results can reasonably be interpreted in a way comparable to that previously employed for achievement motivation. The person scoring high in *n* affiliation is conceived to be positively motivated toward affiliation. When he is unanimously opposed in attitude, he changes his attitude to join his potential friends. When this group of people is divided in attitude, the social reality of the opposing position is much less compelling, as also the perceived necessity to conform to gain social acceptance. Under these conditions, his thinking and outward actions change but little.

The person obtaining a moderate score on *n* Affil is conceived to be ambivalently and anxiously motivated toward affiliation. His fear of rejection leads to a more pronounced conformity. His thinking, as in the case of the positively motivated person, leans heavily on social reality, so that he shifts attitudinally in the nonsupport situation, but much less so under conditions of support. In contrast to the positively motivated person, however, his anxiety for approval tends to maintain his conformity even in the support situation, since he derives less assurance from the presence of a lone partner.

If the low-scoring person experiences a negligible arousal of affiliative motivation in the social-influence situation, then his behavior should be a function of other factors present, in particular the content of the influential communications. To the extent that he is less dependent upon social reality, he should be more free to develop beliefs based on other considerations. Hovland, Lumsdaine, and Sheffield (9) report that for those initially opposed to a point of view being advocated, greater opinion change resulted when both sides of the issue were presented (with the favored view given preponderant consideration) than when only the favored position was considered. This difference was especially pronounced for those of high school education or better. It is apparent that their situation and results are similar to those of the low-affiliation group in the present study, where greater attitude change occurred upon presentation of both sides of the issue, i.e., under conditions of support. The conform-

ity of this group is interpreted as a by-product of attitude change.

Analyses of individual cases, the posttest questionnaire and interview, and the pretest self-ratings give equivocal support to these interpretations, which of course remain highly tentative. Atkinson, in a personal communication, reports data indicating a curvilinear relationship between *n* Ach and *n* Affil, with the high *n*-Affil group also high on *n* Ach, but the medium *n*-Affil group low on *n* Ach. Since McClelland (13) found a strong negative relationship between *n* Ach and independence from group pressures (with subjects from Asch's studies), Atkinson believes the greater conformity of the medium *n*-Affil group in the present study to be due to their lower *n* Ach. It is probable that acceptance of social influence is mediated sometimes by the desire to be "right" or accurate (which is linked to *n* Ach), sometimes by the desire for social acceptance (*n* Affil), and sometimes by other factors. Which factors are involved will depend on the type of issue and the nature of the situation, as well as on the motivational structure of the person. The task set for Asch's subjects, for example, stressed accuracy (cf. 2, p. 452), while the present task did not.

Further research should systematically vary the nature of the task situation and the importance of the issue under study, among other variables, to clarify the relationships involved.

SUMMARY

This investigation studied the effect of affiliative motivation and social support upon conformity and attitude change. College males varying in affiliative motivation were subjected to a group-influence situation where their previously measured attitude toward divorce was either unanimously opposed, or opposed with the exception of one supporter. The subject's public reaction (conformity) was measured, and also his private response (attitude change). The results indicated conformity to be a joint function of affiliative motivation and conditions of social support, with roughly similar but less significant results for attitude change. The group with medium-need scores, interpreted to be ambivalently motivated toward social acceptance, conformed under both conditions, but changed in attitude only in the no-support situation. The high-need group, conceived to be positively motivated toward affiliation, conformed and changed in attitude under condi-

tions of no support, but not under support conditions. The low-need group, which was least affected, changed more under conditions of support than no support, but were considered as responding more to the content of the influence situation than to its social structure.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
3. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405-410.
4. CATELL, R. B. Confirmation and clarification of primary personality factors. *Psychometrika*, 1947, **12**, 197-220.
5. DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.
6. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271-282.
7. FISKE, D. W. Consistency of the factorial structure of personality ratings from different sources. *J. abnorm. soc. Psychol.*, 1949, **44**, 329-344.
8. HEYNS, R. W., VEROFF, J., & ATKINSON, J. W. A manual for scoring thematic apperception stories for the affiliation motive. Unpublished manual, Univer. of Mich., June 1953.
9. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. J. *Experiments on mass communication*. Princeton, N. J.: Princeton Univer. Press, 1949.
10. KRECH, D., & CRUTCHFIELD, R. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
11. LINDQUIST, E. F. *Statistical analysis in educational research*. New York: Houghton Mifflin, 1940.
12. LOWELL, E. L. The effect of need for achievement on learning and performance. *J. Psychol.*, 1952, **33**, 31-40.
13. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
14. MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236-251.
15. MOULTON, R. Relationship of need for achievement to perceptual sensitivity under two degrees of motive arousal. Unpublished honors thesis, Univer. of Mich., 1952.
16. NEWCOMB, T. M. Autistic hostility and social reality. *Hum. Relat.*, 1947, **1**, 69-86.
17. SCHACHTER, S. Deviation, rejection and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190-207.
18. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.
19. SHIPLEY, T. E., JR., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349-356.

TEST-RETEST STABILITY OF INTERACTION PATTERNS DURING INTERVIEWS CONDUCTED ONE WEEK APART¹

GEORGE SASLOW, JOSEPH D. MATARAZZO, JEANNE S. PHILLIPS, AND RUTH G. MATARAZZO
Massachusetts General Hospital and Harvard Medical School²

Two previous studies (5, 6) were devoted to an investigation of the stability of interaction patterns during interviews conducted on the same day by two different interviewers. In the first study (6) of 20 subjects, we reported that interaction patterns for any given individual during a partially standardized psychiatric interview had markedly high stability coefficients across the two interviewers. In addition to demonstrating, for the interview as a whole, the stability of patient interaction patterns (effects) when the stimulus (interviewer's behavior) was controlled along certain dimensions, we observed that the highly stable interviewee patterns could be reliably modified by planned changes in the interviewer's behavior from one sub-period of the interview to another. The second study (5) was an exact replication of the first, and all the major findings were cross validated. Having thus studied stability and predictable modifiability of individual interaction patterns during two interviews conducted on the same day, we decided to examine also the stability of the interaction patterns when the test-retest interval was extended to one week. In order to control for possible interviewer interaction effects (interaction in the statistical sense) the same interviewer was used for both interviews of the present study.

The Interaction Chronograph, devised by Chapple (1, 2) for recording certain temporal aspects of verbal and gestural interaction, was the instrument of measurement used in this and the two previous studies. A more complete description of this method of observation and the variables studied will be found in an earlier report (6). A detailed review of the pertinent literature, and a history of the theory and development of the interaction chronograph, will be found in a recent article (4).

¹ This investigation was supported by a research grant (M-1107) from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

² The data of this study were collected while the authors were at the Washington University School of Medicine.

Briefly, the "rules" of the standardized interview used in this and the two previous investigations require the interviewer to follow certain behavioral patterns during each of five prescribed subperiods of the interview. Periods 1, 3, and 5 consist of free give-and-take interviewing, while Periods 2 (silence) and 4 (interruption) are stress phases of the interview. The interviewer's varied behavior during these five periods constitutes a sequence of different behavioral situations, or what can be thought of as varying miniature interpersonal situations, designed to elicit the interviewee's characteristic interaction patterns. It should be noted that the stress of Periods 2 and 4 is provided by the overt behavior ("not responding" and "interrupting," respectively) of the interviewer, and not by "content" of what he says to the interviewee. The durations of Periods 1, 3, and 5 are fixed (10, 5, and 5 minutes, respectively), whereas the durations of Periods 2 and 4 are variable. The latter periods consist, respectively, of 12 failures to respond or 15 minutes, whichever is shorter; and 12 interruptions or 15 minutes, whichever is shorter.

More complete definitions of the interaction variables can be found in our earlier reports (4, 6). Briefly, they are: (a) *A's Units*, the frequency of the patient's actions; (b) *Action*, the average duration of A's actions; (c) *Silence*, the average duration of A's silences; (d) *Tempo*, the average duration of each action plus its following inaction as a single measure; (e) *Activity*, the average duration of each action minus its following inaction expressed as a single measure; (f) *A's Adjustment*, the average duration of A's "interruptions" of B minus the average duration of A's "failures to respond to" B; (g) *B's Adjustment*, the average duration of the other (usually the interviewer) person's adjustment; (h) *Initiative*, the per cent frequency with which A initiates to B following a double silence, as in Period 2; (i) *Dominance*, the relative frequency with which A dominates B in a double action (interruption), as in Period 4; (j) *A's Synchronization*, the frequency with which A fails to synchronize with B either by failing to

respond to B or by interrupting B; (k) B's *Synchronization*, the frequency of the other person's (usually the interviewer's) failure to synchronize with A; and (l) B's *Units*, the frequency of the other person's (usually the interviewer's) actions.

METHOD

In all aspects but two (the use of only one interviewer, and a seven-day test-retest interval) the present study is an exact replication of the two earlier experiments. Twenty additional outpatient Ss were interviewed twice, seven days apart, by the older of the two interviewers who participated in the previous studies. The rooms and the observer were the same; the Ss were all white; they ranged in age from 16 to 61 (median age of 34.5); and they had presenting complaints typical of the two earlier series of twenty patients each. In the first and second studies there were 11 men and 9 women, and 8 men and 12 women, respectively. The present sex distribution was 9 men and 11 women.

The present series was begun at the same point in time as the first study (6) and was completed at approximately the same point in time as the second study (5). Since the present interviewer was one of the two participants in the earlier studies, it is assumed that the amount of previous experience of the interviewer, himself, with the standardized interview is not a relevant variable in the present study.

RESULTS

As indicated above, twenty minutes of the standardized interview are fixed and, depending upon the subject's behavior in Periods 2 and 4, up to thirty more minutes, totaling fifty in all, are possible to complete the standardized interview. Table 1 shows the mean length of the 20 pairs of interviews for both initial test and retest.

The results indicate that *both* on initial test and retest one week later the interview lasted on the average approximately 32 minutes with the third sample of subjects; a finding similar to the first and second studies. The range, from 25.3 to 41.6 minutes, is almost identical

TABLE 1

ANALYSIS OF MEAN LENGTH OF TOTAL INTERVIEW IN MINUTES FOR GROUP OF 20 PATIENTS INTERVIEWED TWICE BY THE SAME INTERVIEWER

	Mean	Range
Length of all 40 interviews	31.94	25.3 to 41.6
Length of all first interviews ($N = 20$)	31.50	25.3 to 41.6
Length of all second interviews ($N = 20$)	32.40	27.1 to 40.6

(24.5 to 41.1) with that of our second study (5). Also, as in both our earlier studies, the mean lengths of Period 2 (7.40 and 7.72 minutes) and Period 4 (2.37 and 2.34 minutes) which are free to vary were almost identical on both occasions. The test-retest value of rho (.564) for length of Period 2, as in the two earlier studies, is significant at the .01 level; whereas, once more, the extremely limited range in Period 4 makes the obtained value of rho (.180) of little practical use.

Table 2 presents, for the 20 new patients, the means, standard deviations, and ranges (for the total interview) of nine of the interaction chronograph variables.

Mean values for the first and second (repeat) interview are given, with rho and Pearson coefficients of correlation (stability coefficients) shown in the second from the last and last columns, respectively. In these last two columns, rhos and Pearson coefficients of correlation (r) are reproduced from the two earlier studies for comparison. For any given variable the three values in these columns, from top to bottom, represent: the values for the present study, those for the replication study (5), and the values for the original study (6), in that order.

Whether rho or (r) is used, the values in Table 2 show for the present study that, for the interview as a whole, there is again a fairly striking stability in interaction variables for the 20 pairs of interviews separated in time by a seven-day interval. For rho, eight of the nine variables are significant at the .01 level while the remaining one reaches the .05 level. Using r , six coefficients are at the .01 level and three at the .05 level. The ranges for rho and r for these nine variables are .528* to .776**, and .464* to .890**, respectively.

When these are compared to the range of rho for the original (.668** to .874**), and replication studies (.717** to .941**), and the range of r for the original (.726** to .956**), and replication studies (.741** to .926**), it is seen that, despite their statistical significance, the obtained stability coefficients of the present study are slightly lower than those obtained in the two earlier studies with a test-retest interval of a few minutes and the use of two different interviewers.

The variables which seem to have relatively lower stability over a one-week interval are *Dr.'s Adjustment* and *Pt.'s Synchronization*,

TABLE 2

SUMMARY OF DATA FOR MAJOR I-C VARIABLES ACROSS TOTAL INTERVIEW WITH ONE DOCTOR ON TWO OCCASIONS

Variable	One Week Apart		ρ ¹	r ²
	First Interview	Second Interview		
1. Mean Pt.'s Units	72.30	78.30	.765**	.890**
SD	15.75	19.54	(.917**)	(.926**)
Range	43 to 118	48 to 131	(.807**)	(.747**)
2. Mean Pt.'s Action ³	45.70	37.75	.688**	.464*
SD	30.58	16.68	(.914**)	(.786**)
Range	19 to 143	11 to 80	(.757**)	(.956**)
3. Mean Pt.'s Silence	9.15	9.50	.642**	.648**
SD	1.88	2.16	(.828**)	(.744**)
Range	6 to 14	7 to 14	(.731**)	(.759**)
4. Mean Pt.'s Tempo	46.00	44.40	.735**	.764**
SD	14.19	14.59	(.941**)	(.905**)
Range	27 to 96	22 to 84	(.805**)	(.908**)
5. Mean Pt.'s Activity	28.00	25.75	.738**	.768**
SD	14.19	13.60	(.910**)	(.891**)
Range	8 to 76	3 to 62	(.874**)	(.930**)
6. Mean Pt.'s Adjustment	-.89	-1.15	.776**	.757**
SD	.77	1.39	(.821**)	(.853**)
Range	+.58 to -2.50	+.89 to -5.24	(.716**)	(.802**)
7. Mean Dr.'s Adjustment	-2.09	-2.17	.558**	.554*
SD	.77	.87	(.780**)	(.749**)
Range	-.98 to -3.34	-.95 to -3.69	(.668**)	(.737**)
8. Mean Pt.'s Synchron.	.86	.88	.544**	.526*
SD	.05	.07	(.717**)	(.741**)
Range	.76 to .93	.73 to 1.01	(.691**)	(.726**)
9. Mean Dr.'s Units	59.65	66.25	.528*	.823**
SD	13.88	18.49	(.880**)	(.909**)
Range	37 to 96	37 to 118	(.782**)	(.772**)

^{1,2} The three values shown in the Pearson r (and ρ) columns for each variable are, from top to bottom, the r for the present study, the r for replication study, and the r for the original study.

³ Values for variables 2 through 7 are given in hundredths of a minute. To convert to seconds, multiply the given value by 0.6.

* Significant at the .05 level of probability.

** Significant at the .01 level of probability.

both dropping from stability values of approximately .740** in each of the two earlier studies to approximately .540* in the present study. On the other hand, the most stable variable was again found to be *Pt.'s Units*, with a ρ of .765** and an r of .890**.

As can be seen from the means, standard deviations, and ranges in Table 2, all the interaction variables, with the possible exception of *Dr.'s Adjustment* and *Pt.'s Synchronization*, show an absolute as well as relative stability across the two interviews for the interview as a whole. Thus, for example, the ranges for *Pt.'s Units* show that during the same standardized interview one *S* interacted 43 times during the first interview and 48 times during the second interview one week later, whereas a second *S* interacted 118 times during the first interview and 131 times during the second. Similar comparisons can be made for each of the other variables shown in Table 2, although it is apparent from the magnitude of the correlations that, for some variables,

unlike the case just described for *Pt.'s Units*, the extremes of the range need not necessarily represent the same individual in every case.

In all, the data in Table 2 would thus appear to cross validate our earlier finding that, using partially standardized interview conditions, and taking the total interview as the sample, patients show similar interaction patterns from one interview to another. This finding appears to hold whether the test-retest interviews occur on the same day (as in the earlier studies) or one week apart (as in the present study). The similarity on retest is marked in all three studies, but is somewhat greater when the interviews occur only a few minutes apart, even if they are conducted independently by two different interviewers.

Table 3 presents a further cross validation of the earlier finding that these interaction patterns are invariant only as long as stimulus (interview) conditions remain the same across the two interviews for the total interview. Whether on test or retest, as the interviewer

TABLE 3
MEANS AND STANDARD DEVIATIONS FOR MAJOR I-C VARIABLES FOR INDIVIDUAL PERIODS OF TWO INTERVIEWS
WITH ONE DOCTOR ON TWO OCCASIONS ONE WEEK APART

Period		Pt.'s Units		Pt.'s Action		Pt.'s Silence		Pt.'s Tempo		Pt.'s Activity	
		First	Second	First	Second	First	Second	First	Second	First	Second
I	Mean	20.25	24.05	72.80	50.70	8.30	8.25	81.10	59.00	64.70	42.55
	SD	7.28	10.14	110.80	40.85	1.58	2.05	110.30	41.63	117.09	42.52
II	Mean	13.10	13.10	45.55	43.30	14.75	15.00	60.50	58.65	30.95	28.35
	SD	1.76	.44	37.88	25.43	5.88	5.62	40.11	26.46	36.38	25.54
III	Mean	11.00	11.95	53.10	48.85	8.85	8.90	62.20	58.40	44.20	39.90
	SD	4.72	5.22	53.74	28.56	2.71	2.78	52.81	28.84	54.77	28.11
IV	Mean	15.90	15.05	9.35	8.60	5.40	6.40	15.05	15.70	3.95	2.35
	SD	4.54	3.79	4.02	2.38	1.74	3.23	4.13	3.42	4.38	4.98
V	Mean	11.95	14.15	48.05	37.50	8.45	8.60	56.50	46.55	40.00	28.90
	SD	4.27	5.52	22.94	18.03	1.78	2.18	23.66	17.74	22.21	18.72
Total	Mean	72.30	78.30	45.70	37.75	9.15	9.50	46.00	44.40	28.00	25.75
	SD	15.75	19.54	30.58	16.68	1.88	2.16	14.19	14.59	14.19	13.60

Period		Pt.'s Adjustment		Dr.'s Adjustment		Pt.'s Synchronization		Dr.'s Units	
		First	Second	First	Second	First	Second	First	Second
I	Mean	-.91	-1.05	-.63	-.45	.98	.98	18.70	21.75
	SD	1.31	1.33	.75	.21	.06	.05	6.88	9.38
II	Mean	-.97	-.83	-10.87	-11.70	.29	.36	3.75	4.20
	SD	1.05	.79	3.86	3.88	.16	.21	2.16	1.99
III	Mean	-1.38	-1.53	-.67	-.70	1.02	1.01	10.60	11.35
	SD	2.03	1.85	.48	.51	.05	.03	4.60	4.77
IV	Mean	-.67	-1.04	1.22	1.39	.96	1.00	15.10	15.40
	SD	.83	2.48	.90	1.42	.08	.08	3.38	3.06
V	Mean	-.52	-1.30	-.47	-.58	1.00	.97	11.55	13.55
	SD	1.15	1.80	.22	.33	.02	.06	3.99	5.64
Total	Mean	-.89	-1.15	-2.09	-2.17	.86	.88	59.65	66.25
	SD	.77	1.39	.77	.87	.05	.07	13.88	18.49

varied his own behavior within the interview—from one subperiod to another—there is seen in Table 3 a corresponding change in the means obtained by patients with him for any one variable.

Thus, for example, during the first interviews the average interviewee values for silence in Periods 1 to 5, respectively, vary from 8.30 up to 14.75 and back down to 8.85 and down still further to 5.40 and finally up again to 8.45 hundredths of a minute as the interviewer's own silence behavior varies (see "rules" in earlier reports) from responding quickly (within one second of the patient's last utterance) in Period 1 to not responding for 15 seconds in Period 2, to responding within one second again in Period 3, to interrupting the patient in Period 4, and to responding within one second again in Period 5. Also shown in Table 3 is the almost identical pattern for the second, or retest, interview. The value for F across these 5 correlated silence

means (3, pp. 284-296) was 26.06 for the first interview and 26.30 for the second interview a week later. Both of these values are significant at the .001 level of confidence.

Further examination of the data in Table 3 reveals that, as in the two earlier studies, *Pt.'s Action* is also a function of what the interviewer is doing during a particular subperiod of the interview. On the first occasion, for example, when the interviewer was conducting a free give-and-take interview in Periods 1, 3, and 5, the patients' average duration of utterance, *Pt.'s Action*, was 72.80, 53.10, and 48.05, whereas while he was interrupting them in Period 4, the patients' average action fell dramatically to a value of 9.35 hundredths of a minute. Interviewer's lack of action in Period 2 (i.e., silence stress) likewise was associated with a slight drop in patient action to an average value of 45.55. A similar pattern across the five subperiods is evident for patients' action during the repeat, or sec-

ond, interview. The values of F are 10.79 (.001 level) for the second interview and 3.22 (just short of the .01 level) for the first interview.

It should be pointed out that these subperiod means for action and silence, as well as the other variables shown in Table 3, are in reality averages of averages. That is, Chapple obtains an individual's pattern for each subperiod by dividing the obtained value for any given variable by the number of patient units in that period. For example, a patient whose units for the 5 subperiods were 20, 12, 10, 12, and 10, and whose corresponding actions were 800, 288, 400, 144, and 400, would be scored as having average subperiod actions of 40, 24, 40, 12, and 40. For any given individual, these last figures, being averages, are thus relatively independent of both the number of patient units during the 5 different subperiods, and the correlated fact that the 5 periods vary in length (as described earlier). As averages they are also relatively independent of individual differences in interaction rates (patient units) among patients and thus make possible comparisons between individuals. The means shown in Table 3 are therefore averages of these individual averages across the group of twenty patients.³

It is of interest to examine the records of the 20 patients individually to see whether or not the statistically significant group trends shown in Table 3 for *Action* and *Silence* are also true for individual cases. Analysis of the individual records indicates that, independent of each patient's own baseline, all 20 showed a drop in *Action* from Period 3 to Period 4 with a subsequent increase in Period 5. The pattern held individually for all 20 patients in both the first and second interviews. This finding cross validates the earlier finding of a similarly unanimous pattern (5).

As in the previous study, there is a less striking finding of increased *Silence* in Period 2 relative to Periods 1 and 3. In our previous

study (5) 15 individuals out of the group of 20 showed, for both test and retest, a pattern similar to the group pattern shown in the present Table 3; the remaining 5 did not. In the present study 12 out of 20, taken as individuals, showed the same group pattern on both test and retest. The remaining 8 individuals included 6 who showed the group pattern on one occasion but not on the other; and 2 individuals who showed just the opposite pattern, i.e., a decrease in average silence in Period 2 rather than an increase. These 2 "deviates" manifested the same pattern on both test and retest and thus are comparable to the 5 "deviates" of our previous study (5) who likewise were stable in their deviant pattern on the two occasions.

The analysis by individuals as well as the group trends shown in Table 3 are thus a cross validation of our earlier finding that interaction patterns of patients, while invariant under identical inter-interviewer conditions across the interview as a whole (Table 2), are markedly susceptible to planned changes in the stimulus conditions (presented as intra-interview changes from Periods 1 to 5 in the behavior of a single interviewer). Thus Table 3 shows that, as a given interviewer changes his behavior from one period to another, there are predictable associated changes in interviewee interaction patterns. However, Table 2 (as well as Table 3) shows that the same interviewer on retest can elicit almost identical interviewee patterns (invariance) during this interview taken as a whole providing that he presents roughly identical stimulus (or interview) conditions.

In Table 4 can be seen the extent to which the mean patient interaction values shown in Table 3 are reliable from initial test to retest for any one period, in contrast to the interview as a whole. Both rho and r correlation values are shown, although it was demonstrated in our earlier study (5) that the rank-order correlation, rho, is the more appropriate statistic for our data in view of the concomitant problems of extremely narrow ranges, limited frequencies, and some occasional skewing. Values for r are thus included in Table 4 only for purposes of comparison.

In the previous study in which rho was also used, 38 of the total of 45 rhos (9 variables times 5 periods) were significantly reliable; 34 at the .01 level and 4 at the .05 level (5). In

³ The test-retest raw scores for all the interaction chronograph variables of each of the 20 subjects studied here, as well as a description of the method used to obtain them from the printed interaction chronograph record, can be obtained by ordering Document No. 5095 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$2.00 for microfilm or \$3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 4
TEST-RETEST CORRELATIONS FOR EACH OF THE SUBPERIODS OF THE INTERVIEW

Period	Pt.'s Units		Pt.'s Action		Pt.'s Silence		Pt.'s Tempo	
	rho	r	rho	r	rho	r	rho	r
I	.464*	.490*	.450*	.030	.627**	.532*	.368	.023
II	.428*	.052	.653**	.686**	.520*	.592**	.687**	.688**
III	.669**	.665**	.476*	.153	.465*	.331	.487*	.156
IV	.167	.210	.183	.250	.327	.194	-.060	-.186
V	.371	.439	.280	.265	.514*	.500*	.220	.272
Total	.765**	.890**	.688**	.464*	.642**	.648**	.735**	.764**

Period	Pt.'s Activity		Pt.'s Adjust.		Dr.'s Adjust.		Pt.'s Synchron.		Dr.'s Units	
	rho	r	rho	r	rho	r	rho	r	rho	r
I	.518*	.037	.647**	.202	.536*	.550*	.550**	.249	.336	.458*
II	.553**	.679**	.458*	.572**	.535*	.565**	.376	.208	.584**	.580**
III	.429*	.154	.582**	.398	.458*	.293	.617**	-.062	.572**	.670**
IV	.314	.205	.443*	.439	.211	.123	.613**	.240	.323	.447*
V	.235	.248	.588**	.428	.548**	.507*	.671**	.018	.368	.470*
Total	.738**	.768**	.776**	.757**	.558**	.554*	.544**	.526*	.528*	.823**

* Significant at the .05 level of probability.

** Significant at the .01 level of probability.

TABLE 5
MEANS, STANDARD DEVIATIONS, RANGES AND RHOS
FOR INITIATIVE (PERIOD 2) AND DOMINANCE
(PERIOD 4)

	# 1	# 2	rho
1. Mean Initiative (Period 2)	.70	.67	.416
SD	.19	.16	
Range	.29 to .92	.38 to .92	
2. Mean Domi- nance (Period 4)	-.46	-.36	.404
SD	.26	.35	
Range	0 to -.92	+.75 to -.80	

Table 4 of the present study, 29 of the possible 45 period-variable combinations show statistically significant (stable) values of rho; 15 at the .01 level and 14 at the .05 level. Thus, in the subperiods, as well as in the interview as a whole (shown in Table 2), the use of a test-retest interval of seven days, in contrast to several minutes, is associated with somewhat less stability, relatively speaking, for the interaction variables studied.

However, when one remembers the artifacts in measurement described earlier, and also the fact that some of the periods shown in Table 4 are based on only a small fraction of the total (32-minute) interview, the 29 statistically significant rhos out of a possible 45 still indicate considerable stability for the interaction pat-

terns of the 20 patients studied. Thus, for example, the finding of so few statistically significant correlations for Period 4 is probably a function of the fact that this period represents only 2.37 minutes of the total 32-minute interview; the well-known Spearman-Brown phenomenon. Of all the 5 periods shown in Table 4, Periods 3 and 2 elicit, relatively speaking, the most stable values of all nine variables. The stability of Period 3 in this study is in contrast to our earlier finding that this period was one of the least stable (6). However, in our earlier study we had some evidence that this finding was a carry-over of the differences in the way our two interviewers applied the silence stress in Period 2. Since, in the present study, the same interviewer applied the silence stress in both the test and retest of Period 2 this difference in the two experiments is not surprising.

The results of Tables 3 and 4, while somewhat less striking than the comparable findings of our two earlier studies, nevertheless indicate statistically significant (i.e., predictable) modifiability in patient interaction patterns as a function of planned changes in the stimulus conditions presented by the interviewer from one subperiod of the interview to another.

Finally, Table 5 shows the means, standard deviations, ranges, and stability coefficients

for the *Initiative* (Period 2) and *Dominance* (Period 4) variables in the first and second interviews.

The values of rho for both these variables are significant at the .06 level of confidence. These values indicate a fair stability for these interaction variables, in view of the fact that the initiative score is based on Period 2 which was of only about 7.50 minutes' average duration, and the dominance score is based on Period 4 which was of only about 2.37 minutes' average duration. Pearson *rs* for these two variables, .335 and .367, respectively, are slightly lower than the corresponding rhos. Our earlier suggestion (5) would still seem to be pertinent, that the degree of invariance of a single individual's "initiative" or "ascendancy-submission" pattern could be investigated in greater detail (and presumably with a corresponding increase in stability) by extending, in a separate study, the length of each of these two periods to perhaps 15 minutes.

Despite the fair stability (rho) for these two variables, the general lowering of the stability coefficients in the present experiment holds also for both of them since three out of four values for the two earlier studies reached the .01 level and the remaining one (*Dominance* in the original study) reached the .05 level. The .06 level of confidence, found for each of the two variables in the present study, is therefore another indication of a somewhat lowered stability coefficient when the test-retest interval is increased from a few minutes to seven days, even if the same interviewer is used for retest rather than a different interviewer.

DISCUSSION

These results in conjunction with those reported in our earlier studies, seem to indicate two things: (a) a test-retest interval of several minutes and the use of two different interviewers (5, 6), as well as retest interval of seven days and the use of the same interviewer on the two occasions, both result in the finding of considerable stability for the rather complex interaction variables obtained during a psychologically meaningful, miniature interpersonal situation (the interview). It is, however, apparent that (b) the stability is somewhat lower for the second of the two methods employed.

There are undoubtedly many factors associated with this slight lowering of stability in the seven-day study. Four factors, at least,

may be involved: (a) an habituation effect as any given patient "gets used to" the same interviewer on the two occasions; (b) the possibility that the patient's "set" may be different in still another way on the two occasions, e.g., as manifested by a difference in what he talked about (differences in the patient's "content" in the two interviews); (c) the possibility of extra-interview influences in the life of the patient during the seven-day interval, e.g., major life changes such as a decision to leave one's spouse, change jobs, etc.; and (d) a true relative lowering of stability in time simply as a function of "random" or "unknown" life factors, such that one would expect progressively lower stability coefficients as the retest interval was increased to one month, 6 months, one year, etc.

Each of these four possibilities seems amenable to experimental investigation. The "habituation" factor could be explored by employing, for test-retest, as in our two earlier studies, two different interviewers with the same patient (in counterbalanced order) for interviews seven days apart. If stability coefficients higher than those of the present study were obtained the "habituation" hypothesis would be tenable. Our group is currently investigating the hypothesis that the "content" of the two interviews may be more dissimilar in interviews seven days apart than in two interviews conducted on the same day. The third possibility, that important "extra-interview influences" may modify behavior during the seven day test-retest interval, could be investigated by using any of a number of devices (a specially designed questionnaire, an open-ended interview, etc.) at the end of the retest session. The fourth possibility, that of progressive "decay" in stability coefficients as a function of time, could be investigated by using, in counterbalanced order, two interviewers two weeks apart in one study, one month apart in another, 6 months apart in a third, etc. Of course, if the "habituation" factor were also of interest here, one still could employ the latter approach but use only one interviewer for both test and retest.

In contrast to the four factors just discussed, an alternative hypothesis may explain the slight lowering of stability coefficients in the present study. The standard deviations of all nine of the interaction variables shown in Table 2 for the first interview are smaller than

their counterparts in the two earlier studies. (To take one example, while the average number of patient units in all three investigations is approximately 72, the standard deviation of 15 for the present study is considerably lower than the value of 25 found in the two other series.) The smaller standard deviations indicate that the present sample of patients is more homogeneous with respect to all the interaction variables studied. The lowering of stability coefficients from initial test to retest thus may be merely a reflection of the narrower range of variability for each of the nine variables. Choice among the various possibilities mentioned must await further research.

SUMMARY

The present investigation, an extension of two earlier ones, was a study of the interaction patterns of 20 patients who were interviewed by a single examiner in test-retest interviews seven days apart. The results indicate both that patient interaction patterns are remarkably stable, or invariant, over this seven-day interval, and that the interaction patterns can be predictably modified (within either test or

retest interviews) by planned changes in the interviewer's behavior.

REFERENCES

1. CHAPPLE, E. D. The interaction chronograph; its evolution and present application. *Personnel*, 1949, **25**, 295-307.
2. CHAPPLE, E. D. The standard experimental (stress) interview as used in interaction chronograph investigations. *Hum. Organ.* 1953, **12** (2), 23-32.
3. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.
4. MATARAZZO, J. D., SASLOW, G., & MATARAZZO, RUTH G. The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347-367.
5. MATARAZZO, J. D., SASLOW, G., & GUZE, S. B. Stability of interaction patterns during interviews: a replication. *J. consult. Psychol.*, 1956, **20**, 267-274.
6. SASLOW, G., MATARAZZO, J. D., & GUZE, S. B. The stability of interaction chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, 1955, **19**, 417-430.
7. SASLOW, G., GOODRICH, D. W., & STEIN, M. Study of therapist behavior in diagnostic interviews by means of the interaction chronograph. *J. clin. Psychol.*, 1956, **12**, 133-139.

Received May 18, 1956

AUTHORITARIAN ATTITUDES AND PERSONALITY MALADJUSTMENT

ARTHUR R. JENSEN¹

Institute of Psychiatry (Maudsley Hospital), University of London

SINCE the publication of *The Authoritarian Personality* (2), the relationship of ethnic prejudice and authoritarian attitudes to personality and psychopathology has been the subject of a number of studies. Several scales have been used as measures of authoritarian attitudes, and the personality factors underlying authoritarianism have been investigated with an even greater variety of techniques. The most frequently used attitude measures in studies of this type have been the Anti-Semitism (A-S), Ethnocentrism (E), and Fascism (F) scales developed in *The Authoritarian Personality*. The personality side has been assessed by means of clinical interviews (2), psychoanalysis (1), the Thematic Apperception Test (2), the Rorschach (15, 19), the Rosenzweig P-F test, the Allport-Vernon Scale of Values, and other tests (10). The MMPI has been used most often in studying the authoritarian syndrome (2, 3, 8, 10, 11, 12, 17). A survey of these studies, the majority of which have been reviewed by Christie (5), leaves no doubt that personality factors account for a substantial portion of the variance in ethnic prejudice and authoritarian attitudes in general. The question that still remains open, however, concerns the characteristics and the extent of these personality factors.

Two rather different approaches to this question may be discerned. In one approach the emphasis is on describing the psychodynamics (if the investigator is psychoanalytically inclined) or the trait structure (if the investigator is factor-analytically inclined) of the authoritarian personality, but without reference to psychological maladjustment per se. The dynamic approach is exemplified, for example, by the work of Frenkel-Brunswick (see 5, pp. 226-275). The trait structure approach is represented in the work of Eysenck (7), who conceives of the authoritarian personality as representing one extreme of the dimension of "tough-mindedness-tendermindedness," which in turn he conceives as a projection onto the social attitude field of a set of constitution-

ally determined personality variables, viz., introversion-extraversion. In Eysenck's scheme the introversion-extraversion dimension is represented as independent of "neuroticism" or maladjustment. More will be said concerning this theory in the light of the present evidence.

The second, more prevalent, approach has been concerned with the kind and degree of personality maladjustment associated with authoritarian attitudes. Psychiatric diagnostic techniques, particularly the MMPI, have figured prominently in this research. And it is largely at this point that most of the disagreements have arisen. Thus Masling (17) has criticised the idea that authoritarianism is related to psychological maladjustment, and has mentioned investigations on four groups of psychiatric patients that failed to show significant correlation between various measures of authoritarianism (Anti-Semitism, Ethnocentrism, and Fascism scales) and several criteria of psychopathology, including the MMPI. And Maria Levinson (2, p. 968) found that the entire range on the Ethnocentrism scale was represented in a group of psychiatric clinic patients.

The studies which have found more positive evidence of a relationship between maladjustment and authoritarianism (8, 10, 11) have been addressed not to the question of whether psychiatric patients obtain higher or lower scores than "normal" subjects on measures of authoritarianism, but rather to that of whether persons in the normal or nonpsychiatric population who score high on measures of authoritarianism show a greater degree of mental ill-health than persons scoring low on authoritarianism. It seems likely that the relationship between mental ill-health and authoritarianism does not hold in both directions. In other words while it may be possible to have any degree of mental illness without showing authoritarian attitudes, it may not be possible to manifest an extreme degree of authoritarianism without being psychologically maladjusted. One may conceive of the scatter diagram of the correlation between authoritarianism and maladjustment as being more the

¹Public Health Service Research Fellow of the National Institute of Mental Health.

shape of a right triangle rather than the usual ellipse. This rough hypothesis helps to explain some of the seeming contradictions found in the literature, and if it appears at all tenable, as it does to the present writer in viewing the recent research, it would seem that the most satisfactory approach to the problem of authoritarianism and mental health would be to assess the psychological adjustment of individuals at the two extremes on measures of authoritarian attitudes. This essentially was the plan of the present study.

The Prejudice scale. The principal instrument used as a measure of prejudice in this study was the *Pr* (prejudice) scale, which Gough (11) derived by an item analysis of the MMPI, using the Levinson-Sanford Anti-Semitism scale (16) as the criterion. While the *Pr* scale consists of 32 items of the MMPI, it has no more than seven items in common with any other MMPI scale (*Pa*), and one-fourth of the *Pr* items do not appear in any of the other clinical scales.

The reliability and validity of the *Pr* scale were sufficiently high in Gough's studies (11, 12) to warrant its use as an instrument for measuring prejudice on the personality level. The *Pr* scale has been shown to be significantly correlated with anti-Semitic, ethnocentric, and authoritarian attitudes as measured by the A-S, E, and F scales (11). The relative magnitudes of these correlations indicate that the *Pr* scale may be regarded actually as much a measure of authoritarian attitudes as of prejudice. It should also be recognized that the 32 items of the *Pr* scale, constituting about 6 per cent of the total MMPI items, are almost certainly not the only items of the MMPI that are correlated with authoritarianism. Altus and Tafajian (3) subjected the MMPI to an item analysis using high and low scores on the F scale as the criterion, and found 40 items which together correlated .62 with the F scale. Yet only six of the *Pr* items were among these 40. In the present study an item analysis of the *Pr* scale, using *Ss* in the highest and lowest 27 per cent on the F scale as the criterion, showed that 20 of the 32 *Pr* items discriminated between the high and low groups in the F scale at the 5 per cent level of confidence or better. Thus there are at least 66 items of the MMPI which have been shown to be correlated with authoritarian attitudes. It appears, however, that while many of the

MMPI items are related to the authoritarian syndrome, the relationship is tenuous and various items do not have the same degree of discriminatory power in different populations. A likely hypothesis seems to be that the discriminatory power of the items is related to the degree of educational sophistication of the subjects and probably other factors that have educationally selective effects. This possibility is suggested by the fact that 20 of the *Pr* items discriminated between high and low F-scale subjects among the San Diego State College students and only 6 of the *Pr* items were found to discriminate between high and low F-scale subjects in the study by Altus and Tafajian (3), whose subjects were students in a psychology course in the Santa Barbara College of the University of California, and probably represent an educationally, or at least psychologically, more sophisticated group. (Psychology majors in San Diego State College generally had the lowest *Pr* scores of any students in the present study.)

The content of the *Pr* items may be interpreted in general terms as representing attitudes of anti-intellectuality ("I like science"—*false*, "I like poetry"—*false*), cynicism ("Most people make friends because friends are likely to be useful to them"—*true*), distrust ("I think most people would lie to get ahead"—*true*), doubt and suspicion ("I commonly wonder what hidden reason another person may have for doing something nice for me"—*true*), misanthropy and querulousness ("The man who provides temptation by leaving valuable property unprotected is about as much to blame for its theft as the one who steals it"—*true*), discontent with self-status and fearful self-concern ("It makes me feel like a failure when I hear of the success of someone I know well"—*true*). "Several times a week I feel as if something dreadful is about to happen"—*true*).

The fact that certain MMPI items differentiate between high and low scorers on measures of prejudice and authoritarianism and that the *Pr* scale is correlated with the A-S, E, and F scales has led some writers (3, 11) to characterize the authoritarian personality in terms of the content of these MMPI items. Strictly speaking this is not a valid procedure. In order to be valid, an additional investigation would be required to determine whether or not the items were factually true statements about the individuals taking the test. Since this has

never been done, and indeed would not even be advisable, we can legitimately make no use of item content other than as a possible source of hypotheses about personality.

A more defensible approach would consist of determining the relationships between the authoritarian measures and the various clinical scales of the MMPI and then characterizing the authoritarian personality in terms of our empirical knowledge of the meaning of the clinical scales, which have been elucidated by innumerable studies (6). Even then, caution must be taken not to think of the authoritarian personality in the singular. The approach suggested above would permit a number of generalizations about the personality characteristics to be found among a group of persons having authoritarian attitudes, but most likely only some of these generalizations would hold true for any individual.

METHOD

Subjects. The group form of the MMPI was administered to 712 entering freshmen in the San Diego State College, a four-year coeducational liberal arts college in Southern California, and was readministered one year later to the 312 students of this class who remained in college to begin their sophomore year. The same data and in addition the California F scale (Form 60A) were obtained from 114 seniors majoring in education. Intelligence test (American Council on Education) scores and grade-point averages were also available for all the freshmen.²

The one rather atypical feature in this sample seems to be the high drop-out rate (56 per cent) between the freshman and sophomore years. The writer can only add his speculations to the few points in the data that suggest possible causes for so many drop-outs. For one thing, the "G.I. Bill" was still in effect at the time these data were gathered and large numbers of men were caused to enter college who otherwise would not have sought college training, not only for lack of financial means but also of the necessary academic interests and aptitudes. The college entrance requirements were relaxed for these students, while academic standards were maintained at a high level. Thus many were discouraged from continuing beyond their first year. The mean *Pr* score of the students who dropped out of college was significantly ($p < .05$) higher than those who continued, suggesting that the drop-outs were more like noncollege groups and high school students who do not plan to go to college (13, p. 265). These groups typically are known to obtain higher *Pr* scores on the average than college groups.³ Another cause of

TABLE 1
T CONVERSION TABLE AND PERCENTILES FOR *Pr*
SCORES, BASED ON 712 COLLEGE FRESHMEN
(342 MALES, 370 FEMALES)

Raw Score	f	Percentile	T Score*	Raw Score	f	Percentile	T Score
28	1	99	**	13	23	83	60
27			77	12	43	79	58
26			76	11	40	73	56
25			76	10	50	67	55
24			75	9	41	60	53
23	1	99	75	8	66	53	51
22			74	7	62	44	48
21			73	6	64	35	46
20			71	5	55	27	44
19			70	4	56	19	41
18	13	97	69	3	48	12	38
17	14	95	67	2	41	5	34
16	14	93	65	1	11	2	29
15	24	90	63	0	7	1	24
14	27	87	61				

* Since the distribution of raw scores is significantly skewed, the *T* scores have been normalized.

** The absence of frequencies above this point prohibits the computation of *T* scores.

the high drop-out rate is that for many students the state college is a stepping stone to other colleges and universities.

RESULTS

Characteristics of Pr. Since normative data on the *Pr* scale have been published only for a sample of high school seniors (11, 12), it would seem advisable to present similar data for a college sample before going on to make further interpretations on the basis of *Pr*.

The distribution of raw *Pr* scores in the college sample is conspicuously more constricted and positively skewed than in Gough's high school sample. In order to make comparisons of the *Pr* scale with other MMPI scales, a *T* conversion table was prepared (Table 1). This *T* scale has a mean of 50 and a standard deviation of 10. Since a test of the skewness of the distribution of raw scores was significant at the 5 per cent level, the *T* scale was normalized.

Table 2 is quite self-explanatory. The following points in this table are to be noted.

a. The reliability of *Pr* is consistent with Gough's findings (11, p. 254) and compares favorably with the reliability of other MMPI scales (see 6, p. 5).

b. The retest reliability after one year is notably high considering the restricted range of scores.

c. The mean *Pr* for the freshmen is approximately one standard deviation lower than that

² I am indebted to Dr. Kenneth Eells for having made these data available to me.

³ The writer also administered the *Pr* and *F* scales to 96 junior college vocational students in the same locality and found their mean scores on both scales to be significantly ($p < .001$) higher than those in the liberal arts college.

TABLE 2
DATA ON *Pr* SCALE IN TWO COLLEGE SAMPLES

<i>Pr</i> Data	College Freshmen <i>N</i> = 712	Education Majors (Seniors) <i>N</i> = 114
a Split-half reliability (odd-even, corrected by Spearman-Brown formula)	.81 (<i>N</i> = 100)	
b Retest reliability (1 year interval)	.56 (<i>N</i> = 312)	
c Mean and Standard Deviation	8.35 <i>SD</i> 4.59	6.37 <i>SD</i> 3.43
d Average decrease in <i>Pr</i> after 1 year of college and Critical Ratio for the difference	2.05 <i>CR</i> 9.45 (<i>N</i> = 312)	
e Correlation of <i>Pr</i> with <i>F</i> scale (Form 60A)*		.27 <i>SE</i> .09
f Correlation of <i>Pr</i> with intelligence (American Council on Education test)	-.15, <i>SE</i> .06 (<i>N</i> = 288)	
g Correlation of <i>Pr</i> with Grade Point Average	-.13, <i>SE</i> .06 (<i>N</i> = 288)	

* This scale is identical with the E-F scale given in Gough paper (10, pp. 239-241).

of Gough's high school sample, and the mean of the education majors is nearly two standard deviations lower. This is consistent with the findings of other studies, which report significant correlation between educational level and measures of authoritarianism (5, p. 170).

d. The systematic lowering of *Pr* scores after one year of college is highly significant. While 69 per cent of the *Ss* obtained lower *Pr* scores on the retest, only 20 per cent obtained higher scores. The 11 per cent whose scores did not change over the one-year interval had a very low mean *Pr* of 5.77, *SD* 3.71.

e. The correlation of *Pr* with the *F* scale in the college senior group is considerably lower than the correlation of .46 in Gough's study. This is undoubtedly due to the very restricted range of scores on both scales in the college sample, a factor tending to lower the correlation coefficient. When the present sample was combined with a group of 96 junior college vocational students who had taken the same tests, the correlation between *Pr* and *F* rose to .65, *SE* .07.

f & g. The correlations of *Pr* with intelligence and grade-point average are similarly

much lower than those in Gough's study, again because of the greater homogeneity of the college sample, with the consequent restriction of range on the correlated variables. Both the correlations, however, are statistically significant at better than the 5 per cent level of confidence. While there is certainly a negative correlation between intelligence and *Pr*, as well as of other measures of authoritarianism (5, p. 168), intelligence may be regarded as a negligible factor in the variance of *Pr* in a college population. For example, it was found that groups even with widely differing *Pr* scores (above 15 and below 3) showed no significant difference in mean intelligence test (ACE) scores.

Some additional findings about *Pr* may be noted. There were no significant differences between men and women. Students majoring in different subjects had significantly different mean *Pr* scores. Why these differences should exist among groups of students of the same age, at the same stage in their education, and in the same college would be difficult to explain except in terms of personality differences. Most of the groups representing different fields of study were too small to permit meaningful statistical comparisons, but the two largest groups—business majors and education majors—showed a highly significant difference ($p < .001$) in mean *Pr* scores. The high *Pr* groups were students majoring in business administration, commerce, social work, art, music, mathematics, and chemistry. The low *Pr* groups were students majoring in psychology, education, physics, biology, and economics. Any generalizations or interpretations from these data alone would necessarily be *ad hoc*. Since the students in all these groups took the MMPI at the very beginning of their freshman year, there can be no reflection on the influence of the faculties or curricula of the various departments of the college in contributing to these differences in mean *Pr*.

Pr and personality adjustment. The greatly restricted range of the *Pr* distribution in this college sample, with the large piling-up of scores toward the low end of the scale, and the likely constriction and skewness of the distribution on most of the clinical scales of the MMPI, deterred the investigator from studying the relationship between *Pr* and the clinical scales by means of the correlation co-

efficient. The nature of the distributions in this sample would have so attenuated the correlations as to obscure possibly important relationships. Therefore it was decided to compare the MMPI clinical scales of adequately large groups of high and low *Ss* on *Pr* to permit the emergence of relatively unambiguous statistical significance should real differences exist. While it is the standard practice to use the upper and lower 27 per cent of the distribution in comparing high and low groups, this cutoff point would be impossible in the present distribution, as is obvious from inspection of the Percentile column in Table 1. There would be only five points difference between the highs and lows if the upper and lower 27 per cent were used, and furthermore the highs would still fall below the mean *Pr* of noncollege groups. Therefore it was decided to select from the 712 freshmen who took the MMPI the upper 7 per cent ($N = 54$) on the *Pr* scale (scores above 15) and the lower 8 per cent (scores below 3, $N = 59$).

Comparisons between these groups on the

nine clinical scales and the *K* scale of the MMPI revealed that some of the scales discriminated positively, some negatively, and others not at all. On each scale that showed a difference between the high and low *Pr* groups, this difference was significant at better than the .001 level of confidence. The differences on the remaining scales were in all cases smaller than the standard error of the difference and thus completely nonsignificant. The high *Pr* group was higher (more "maladjusted") on the following MMPI scales: *Hs* (hypochondriasis), *D* (depression), *Pd* (psychopathic deviate), *Pt* (psychasthenia), *Sc* (Schizophrenia), and *Ma* (hypomania). The following scales discriminated negatively: *K* ("defensiveness") and *Hy* (hysteria). The *Mf* (masculinity-femininity) and *Pa* (paranoia) scales did not discriminate significantly. These results are presented in Table 3 in such a way as to permit comparisons with other studies on the relationship between measures of authoritarian attitudes and the MMPI scales. The magnitude of the correlations has not been given, as

TABLE 3

RELATIONSHIPS¹ BETWEEN VARIOUS MEASURES OF AUTHORITARIANISM AND THE MMPI SCALES

Study	Sample	N	Measure of Relationship	Criterion Scale	MMPI Scales										
					<i>K</i>	<i>Hs</i>	<i>D</i>	<i>Hy</i>	<i>Pd</i>	<i>Mf</i> (m)	<i>Mf</i> (f)	<i>Pa</i>	<i>Pt</i>	<i>Sc</i>	<i>Ma</i>
Jensen	College Freshmen	113	r^2	<i>Pr</i>	—	+	+	—	+			—	+	+	+
Tyler (20, p. 453)	Graduate Education Majors (female)	107	r	<i>Pr</i>	0 ³		+	—		+		+	+	+	+
Gough (12, p. 258)	High School Seniors ⁹ Class A	271	r	<i>Pr</i>	—		+		+		+		+	+	+
Gough (12, p. 258)	High School Seniors Class B	231	r	<i>Pr</i>	—		+		+	—	—			+	+
Gough (10, p. 241)	High School Seniors Class A	271	r	A-S ⁴	—		+		+					+	
Gough (11, p. 249)	High School Seniors Class A	80	r^5	A-S ⁴	—		+		+						
Gough (10, p. 241)	High School Seniors Class A	271	r	E-F ⁶	—			—		—					
Freedman, et al. (8, p. 317)	College Freshmen Sample 1 (female)	441	r	F ⁷	—	+		— ⁸				—	+		—
Freedman, et al. (8, p. 317)	College Freshmen Sample 2 (female)	225	r	F ⁷	—			— ⁸							

¹ + indicates a positive correlation, — indicates a negative correlation. Only relationships significant beyond the .05 level of confidence are given. In the Jensen study all relationships are significant beyond the .001 level.

² t test for significance of differences between means of highest 7 per cent (*Pr* scores above 15) and lowest 8 per cent (*Pr* scores below 3) in a sample of 712 *Ss*.

³ *K* scale not used in Tyler's study.

⁴ Levinson-Sanford Anti-Semitism Scale (16).

⁵ t test for significance of differences between means of highest 40 and lowest 40 on A-S scale in a sample of 271 *Ss*.

⁶ Gough's *E-F* scale (10, pp. 239-241) is made up of 30 items from the California F scale, Form 60 (2, pp. 248-250).

⁷ California F scale (2, pp. 255-257).

⁸ The somatic items of the *Hy* scale were not counted for this correlation in Freedman's study. The correlation for the complete *Hy* scale was non-significant.

⁹ The Class A High School Seniors in Gough's study is the same sample in every instance; therefore the four sets of data from Gough are not independent.

these vary considerably from one sample to another and essentially add nothing to the picture. Only those relationships significant beyond the 5 per cent level have been included.

Though these results clearly establish the fact of a relationship between *Pr* and certain clinical scales, they alone are not sufficient to answer the question concerning the relation of *Pr* to psychological maladjustment. The correlation, or differences between high and low groups, could conceivably be based entirely on MMPI scores considered to be well within the normal range of psychological adjustment. In order to answer this question the high and low *Pr* groups were compared with respect to the proportion of each group obtaining *T* scores above 70, i.e., the level generally indicative of psychological maladjustment. The results were clear-cut. A significantly ($p < .01$) greater proportion of the high *Pr* group obtained *T* scores above 70 on the following scales: *D*, *Pd*, *Pa*, *Pt*, *Sc*, *Ma*. For each of these scales, on the average, there were about twelve times as many of the high *Pr* group who obtained *T* scores above 70 as of the low *Pr* group. The scales on which the proportions of high and low *Pr* groups having *T* scores above 70 were not significantly different were *Hy*, *Hs*, and *Mf*.⁴ It should be noted that while in the comparisons of Table 3 the *Hy* scale is inversely related to *Pr*, the high and low *Pr* groups do not differ in the proportions obtaining *T* scores above 70 on the *Hy* scale. This finding suggests that though low *Pr* is associated with high *Hy* in this sample, the *Hy* is not so high as to be in the abnormal range. The nonhomogeneous character of the *Hy* scale also enters the picture here. For example, Freedman *et al.* (8) found a significant negative correlation between *Hy* and the California F scale in their college sample only when they removed the somatic items from the *Hy* scale.

Another question remains to be answered concerning the relationship of *Pr* to adjustment. That is, do students who show signs of poor psychological adjustment as judged independently of the MMPI also have higher *Pr* scores than students who appear to be well-adjusted? To get at least a rough answer

to this question the entire college faculty was asked to submit the names of students whom they knew personally and considered either well-adjusted or poorly adjusted in the sense of being likely candidates for psychological counseling. None of the faculty had any knowledge of the aims of this study. Students about whom there was general agreement by two or more of the faculty were compared on the *Pr* scale. Again the results were clear-cut. The poorly adjusted group ($N = 30$) had a significantly ($p < .001$) higher mean *Pr* (9.93) than the well-adjusted group ($N = 57$) with a mean *Pr* of 5.61.

DISCUSSION

There is quite clearly a relationship between authoritarian attitudes on the personality level as measured by the *Pr* scale and psychological maladjustment as measured by the MMPI as well as by faculty ratings. But in addition to this general finding, what more specifically can be gleaned from these data as to the aspects of personality tapped by the *Pr* scale? While in any one study in which a number of correlations or group differences are obtained certain of these may lack general significance because of peculiarities of the particular sample, the type of analysis applied, or the merely chance fluctuations of sampling, the marked consistencies shown in the studies summarized in Table 3 do provide a sound basis for interpretation. The *K*, *D*, *Hy*, *Pt*, and *Sc* scales display the most consistent relationship to measures of authoritarianism, and so the discussion is confined to these.

Not only does *K* show negative correlations with authoritarian measures in every study, but these correlations are generally higher than any of the others. Gough (12), for instance, found correlations between *Pr* and *K* which were nearly as high as the reliabilities of either of these scales. The meaning of *K* is complex. The scale was originally devised as a correction factor for the clinical scales (18) but has since been found to have psychological significance in its own right. In the clinical interpretation of the MMPI a high *K* score is regarded as indicative of a defensive attitude, and a low *K* score suggests unguarded and indiscrete frankness and a tendency to speak critically of self. Low *K* is found most frequently in conditions characterized by ego

⁴ The *K* scale was not included in these comparisons since it has not been validated as a measure of maladjustment.

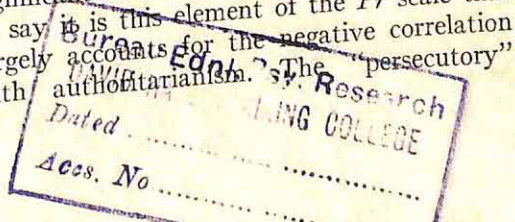
weakness, in which the person is psychologically naked and vulnerable, as in the psychoses and anxiety states. *K* shows a high negative correlation ($-.74$) with the Taylor Manifest Anxiety Scale (4), and a factor analysis of the MMPI in a college sample by Wheeler (21) shows a low *K*, along with high *Pt* and *Sc*, to be highly saturated with the factor interpreted as anxious self-concern with which the ego defenses seem inadequate to cope. High *K*, along with high *Hy*, was saturated with the factor interpreted as intactness of the ego-defense mechanisms. Wheeler's study also highlights the inverse similarity between *K* and *Pr*, in that correlations between *K* and the other MMPI scales show very much the same pattern as does the *Pr* scale, except of course that the direction of the correlation is reversed, since *Pr* and *K* are negatively correlated. High *K* scores have been found to be associated with the following traits: sociable, wide interests, reasonably enthusiastic, and verbal; low *K* scores: high-strung, cynical, dissatisfied, and individualistic (14, p. 78). *K* is positively correlated with intelligence and socioeconomic status and there is a tendency for college students and college-educated persons to obtain *K* scores one-half to one standard deviation higher than the noncollege population—the same relationships that have been found (inversely of course) for the *Pr* scale (6, 15). Of all the MMPI scales, *K* is probably the best single index of the personality factor tapped by *Pr*, at least in the nonpsychiatric population.

The positive correlation between ethnic prejudice and the *D* scale was first noted in *The Authoritarian Personality*, somewhat to the surprise of the investigators, since clinical signs of depression were more apparent in psychiatric patients scoring low on ethnocentrism. As seen in Table 3, however, the positive relationship of authoritarian attitudes with *D* is entirely consistent in several studies. That *D* is not a homogeneous scale and must always be interpreted in relation to other scales in clinical practice may account for the apparent discrepancies with the clinical observation noted in *The Authoritarian Personality*. There is considerable overlap in the traits associated with high *D* and low *K*. Persons obtaining high *D* scores are characterized by lack of self-confidence, tendency to worry,

narrowness of interests, poor morale, an uneasy self-concern and dissatisfaction with their current situation (6). Except in relation to other scales *D* is difficult to interpret clinically. However, taken alone it may be regarded as the best single index of maladjustment in the MMPI (9).

The *Hy* scale shows a consistently negative correlation with the *Pr* and *F* scales. Since *Hy* and *K* are highly correlated, a similar interpretation as that given for *K* applies also to *Hy*: in short, there appears to be a negative relationship between repressive tendencies and authoritarianism. Like *K*, *Hy* is positively correlated with intelligence and is usually about one-half standard deviation higher in college samples than in the adult standardization group. It is quite clearly the psychic and not the somatic items of the *Hy* scale that account for its negative correlations with both *Pr* and *F*. Freedman *et al.* (8) found the correlations between the *Hy* scale and the California *F* scale in two college samples to be $-.02$ and $.00$. But when the somatic items were removed, the correlations rose to $-.41$ and $-.44$. A high *Hy* score based on the psychic items indicates a tendency not to tire easily, not to be depressed, a feeling that life is good and that the world is benevolent—in short, a rather optimistic, extraverted attitude.

The negative correlation (or absence of correlation) between authoritarianism and the *Pa* (paranoia) scale is usually viewed with surprise, for *The Authoritarian Personality* called upon the paranoid mechanisms of projection, extrapunitive, and power orientation in explaining the authoritarian syndrome. The reason for the negative correlation between *Pa* and *F* (and presumably also for the lack of correlation with *Pr* in the present study) has been adequately explained by Freedman *et al.* (8) in terms of the heterogeneous nature of the *Pa* items. When the *Pa* scale is broken down into three types of items, the correlations between *F* and these sub-*Pa* scales are as follows: Persecutory Ideas ($r = .01$), Poignancy ($r = -.07$), and Naïveté ($r = -.22$). This last correlation was significant at the .001 level and it seems safe to say it is this element of the *Pr* scale that largely accounts for the negative correlation with authoritarianism. The "persecutory"



items are not at all subtle, hardly concealing their pathological implications, and comparatively few of them are responded to in the keyed direction by more sophisticated Ss such as college students. But as the *Pa* score rises, more of the "persecutory" items are of course included, and it should be noted that when the high and low *Pr* groups in the present study were compared for the proportions of Ss obtaining *T* scores over 70, the high *Pr* group was significantly *higher* on *Pa*, because more of them responded in the keyed direction to a larger number of *Pa* items, including the "persecutory" items, which are most commonly associated with paranoia.

The *Pt* scale is a relatively pure measure of neuroticism of the anxiety and obsessive-compulsive varieties. Persons scoring high on *Pt* show excessive doubts, compulsions, obsessions or unreasonable fears, anxiety, perplexity, and apprehension.

The *Sc* scale in nonpsychiatric populations is associated with schizoid tendencies, feelings of isolation, disinterest, and pessimism.

In the over-all picture presented by these scales—those most consistently related to authoritarian measures—the emphasis is on the obsessive-compulsive syndrome (high *D*, *Pt*, *Sc*) and there is comparatively little hysterical and repressive tendency (low *Hy* and *K*). The impression is that prejudiced, authoritarian persons have less well-developed ego defenses and are thus more exposed and vulnerable to psychological stress, in the face of which they develop tendencies toward pessimism, cynicism, low morale (*D*), and psychological isolation (*Sc*), along with the more primitive defenses of a compulsive, ritualistic, and schizoid nature (the triad *D*, *Pt*, *Sc*).

In view of the picture presented by the data of this study, as well as the consistencies seen in several others, it seems to this writer that an adequate theory of social attitudes, particularly those in the domain of ethnic prejudice and authoritarianism, must take into account psychopathological aspects of personality. To ignore them seems an unwarranted oversimplification of this complex field. Eysenck, for example, has formulated a theory of social attitudes which essentially leaves out of account the relationship between

social attitudes and personality adjustment (7). Eysenck conceives of social attitudes as being determined by two orthogonal factors—the Radicalism–Conservatism dimension (*R*-factor) and the Toughminded–Tenderminded (*T*-factor) continuum. We shall be concerned here only with the *T* factor, which is regarded as a projection onto the social attitude field of a fundamental dimension of personality, viz., introversion–extraversion which is said to be constitutionally determined and is conceived as being entirely unrelated to emotional instability or neuroticism (7, p. 179). Eysenck equates the authoritarian–democratic continuum with the *T* factor, authoritarian attitudes being an expression of toughmindedness. Furthermore he considers the *A-S*, *E*, *F*, and *Pr* scales as essentially measures of toughmindedness (7, pp. 148, 152, 233). They should therefore be highly correlated with measures of extraversion and show comparatively little, if any, correlation with measures of neuroticism. Eysenck has also stated that those toward the extraverted end of the continuum develop symptoms of the hysterical type during periods of emotional instability, while those toward the introverted end develop "dysthymic" disorders (anxiety, depression, and obsessive-compulsive).

If the writer understands Eysenck correctly, these relationships postulated by his theory are not in accord with the facts presented here. The measures of authoritarianism were shown to correlate negatively with the *Hy* (hysteria) scale and positively with the "dysthymic" scales (*D*, *Pt*), both correlations being contrary to predictions from Eysenck's theory. Two factor-analytic studies provide similar contradictory evidence as well as demonstrate that measures of authoritarianism have a considerable loading on the neuroticism factor. Tyler (20) performed a centroid factor analysis of fifteen MMPI scales, including *Pr*, on a college sample and found that *Pr* had its most substantial loading (.70) on the first factor, which he interpreted as "general maladjustment." The scale showing the highest correlation with *Pr* was *Pt*, which Eysenck considers to be a relatively pure measure of neuroticism. A Lawley-type factor analysis was carried out in this laboratory by Dr. M. I. Soueif on several MMPI scales (*K*,

Hy, *Pd*, *Pt*) and Guilford's factorially pure *D* (depression), *C* (cycloid) and *R* (rhythymia scales. (The *R* scale is regarded by Eysenck as one of the best measures of extraversion.) The first two factors in Soueif's analysis were neuroticism and introversion-extraversion. The MMPI scales with the largest factor loadings on neuroticism were *K* (-.64) and *Pt* (.85), both of which are highly correlated with authoritarianism. The *K* and *Pt* scales had comparatively small factor loadings on introversion-extraversion (.40 and .14 respectively). The *Hy* scale had a loading of .12 on the neuroticism factor and .77 on extraversion. But it has consistently shown negative correlations with measures of authoritarianism, which is exactly the opposite of what one would predict from Eysenck's theory.

SUMMARY

The MMPI was administered twice, with a one-year interval, to 712 college freshmen and to 114 seniors, who also took the California F (Fascism) scale. A discussion of the nature of the *Pr* (prejudice) scale of the MMPI and its relation to a number of other variables in a college sample are presented, along with a normalized *T* conversion table for *Pr*.

1. The distribution of *Pr* scores in the college sample was more skewed and constricted, with a piling up of low scores, than the distribution of scores in the high school group used in the derivation of the *Pr* scale.

2. The reliability of *Pr* (.81) compares favorably with that of other MMPI scales.

3. There was a significant decrease in mean *Pr* as students advance in college.

4. There were significant differences in mean *Pr* among students majoring in different fields.

5. *Pr* correlated significantly with the California F scale.

6. *Pr* was positively correlated with the *Hs*, *D*, *Pd*, *Sc*, and *Ma* scales, and negatively correlated with *K* and *Hy*.

These findings and those of several related studies are discussed as a means of psychologically characterizing the authoritarian personality. It was concluded that certain psychopathological factors are present in the authoritarian syndrome.

REFERENCES

1. ACKERMAN, N. W., & JAHODA, M. *Anti-Semitism and emotional disorder*. New York: Harper, 1950.
2. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
3. ALTUS, W. D., & TAFEJIAN, T. T. MMPI correlates of the California E-F Scale. *J. soc. Psychol.*, 1953, **38**, 145-149.
4. BRACKBILL, G., & LITTLE, K. B. MMPI correlates of the Taylor scale of manifest anxiety. *J. consult. Psychol.*, 1954, **18**, 433-436.
5. CHRISTIE, R., & JAHODA, MARIE (Eds.). *Studies in the scope and method of "The authoritarian personality"*. Glencoe, Ill.: Free Press, 1954.
6. COTTLE, W. C. *The MMPI, A review*. Lawrence, Kansas: Univer. of Kansas Pub., 1953.
7. EYSENCK, H. J. *The psychology of politics*. New York: Praeger, 1954.
8. FREEDMAN, M., WEBSTER, H., & SANFORD, N. A study of authoritarianism and psychopathology. *J. Psychol.*, 1956, **41**, 315-322.
9. GOUGH, H. G. A new dimension of status: II. Relationship of the *St* scale to other variables. *Amer. sociol. Rev.*, 1948, **13**, 534-537.
10. GOUGH, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 237-246.
11. GOUGH, H. G. Studies of social intolerance: II. A personality scale for anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 247-255.
12. GOUGH, H. G. Studies of social intolerance: III. Relationship of the *Pr* scale to other variables. *J. soc. Psychol.*, 1951, **33**, 257-262.
13. GOUGH, H. G. Studies of social intolerance: IV. Related social attitudes. *J. soc. Psychol.*, 1951, **33**, 263-269.
14. HATHAWAY, S. R., & MEEHL, P. E. The Minnesota Multiphasic Personality Inventory. In *Military clinical psychology*, TM 8-242, AFM 160-45. Washington: U. S. Government Printing Office, 1951.
15. HIMMELHOCK, J. The dynamics of tolerance. Unpublished doctor's dissertation, Columbia Univer., 1952.
16. LEVINSON, D. J., & SANFORD, R. N. A scale for the measurement of anti-Semitism. *J. Psychol.*, 1944, **17**, 339-370.
17. MASLING, J. M. How neurotic is the authoritarian? *Amer. Psychologist*, 1953, **8**, 402-403. (Abstract)
18. MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The Minnesota Multiphasic Personality Inventory: VI. The *K* scale. *J. consult. Psychol.*, 1948, **12**, 20-31.
19. REICHARD, SUZANNE. Rorschach study of prejudiced personality. *Amer. J. Orthopsychiat.*, 1948, **18**, 280-286.
20. TYLER, F. T. A factorial analysis of fifteen MMPI scales. *J. consult. Psychol.*, 1951, **15**, 451-456.
21. WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134-141.

Received May 28, 1956

FACTORS DETERMINING THE GUESSING OF MEANINGS OF FOREIGN WORDS¹

YVONNE BRACKBILL AND KENNETH B. LITTLE

Stanford University

IN A recent article, Brown, Black, and Horowitz (1) report an experiment in which 86 English-speaking subjects were asked to guess the English meanings of 21 pairs of antonyms presented in three foreign languages—Hindi, Czech, and Chinese. Inter-subject agreement as to meaning of the foreign words ran as high as 97 per cent and, furthermore, the number of correct guesses for each language was significantly greater than chance expectation. The authors interpret their results as supporting the hypothesis that there is some sort of intrinsic correspondence between the sounds of certain words and their meanings regardless of language. They suggest that, "It is possible . . . that the origin of speech in gestural imitation and physiognomic language has left a residue of roughly translatable words in all languages. . .," and conclude, "For the present we prefer to interpret our results as indicative of a primitive phonetic symbolism deriving from the origin of speech in some kind of imitative or physiognomic linkage of sound and meanings."

The results presented by Brown, Black, and Horowitz are unequivocal in that subjects certainly guessed the correct English equivalents of the foreign words with a surprising degree of accuracy. There are certain aspects of the study, however, that make it difficult to determine whether these results support an explanatory hypothesis of a universal sound-meaning linkage:

1. The method used by the authors was not a fully appropriate test of the hypothesis. Two points may be made in this regard. First, Brown, Black, and Horowitz used visual as well as auditory stimuli in their method of presentation. However, the hypothesis concerns a correspondence between meaning and *sound*, not between meaning and the written form of a word. Therefore, the method most appropriate to a test of the hypothesis would be one depending exclusively on auditory cues.

Second, the hypothesis is necessarily concerned with meanings that have some intercultural communality. It does not apply to associational meanings of the type that are largely unshared, i.e., private or idiosyncratic, or that are shared only by a very narrow segment of the population. With English-speaking subjects, the most adequate methodological control for association meanings lies in avoiding the use of English words, using instead a method that pairs one foreign sound against a second foreign sound, with no meaning provided.

2. The authors made no check on the fidelity of their foreign translations. Invariably there are concepts in one language that have no exact and/or single equivalent in another (cf. 4, p. 191). When a translator must choose that particular word which best approximates the intended meaning, it is quite possible that his choice may simply be in error or that he may select that alternative which most closely resembles the English word in length and sound. Hence, it is desirable to have some consensus of translators as well as a back-translation.

3. The population of concepts is indeterminate and the sampling procedures are ambiguous. Brown, Black, and Horowitz state, "English antonyms were selected from the Thorndike-Lorge word list with two considerations in mind. (a) The words should name sense experiences (e.g., warm-cool, heavy-light). (b) The members of a pair should both fall in the frequency range of 100 or over per million. The final list of 21 pairs consisted of those on which the three authors could agree" (1, p. 389). Yet there are certain discrepancies between their criteria and their actual word sample: (a) Thirteen of the forty-two words have frequencies of use less than 100 per million, by the Thorndike-Lorge word list; (b) three of the word pairs are clearly not antonyms (thunder-lightning, iron-gold, and

¹ This study was supported in part by a Stanford University Research Grant.

many-one); and (c) the sense experience character of the words is quite debatable.²

4. Finally, the authors suggest that some universal sound-meaning relation exists in *all* languages. However, those used in their study and in the unpublished studies they cite constitute a small and unrepresentative sample of the population of existing languages. Obviously there are practical limitations to securing a representative language sample. Possibly the best approximation lies in structuring the sample so that it contains not more than one member of any language group. The sample languages, that is, should be non-cognate. Unfortunately, three of the four languages used by Brown, Black, and Horowitz (English, Czech, and Hindi) *are* members of the same language group.

Maltzman, Morrisett, and Brooks (3) have reported a series of experiments similar to that of Brown, Black, and Horowitz. In their study, subjects were presented, in printed form only, with a single stimulus word and two response words. The languages used were English, Japanese, and Croatian rotated in position as stimulus and response. The sample of concepts used was that of Tsuru and Fries (5) which is a mixture of antonyms and contradictory word pairs. Their results, in English-Foreign comparisons, were similar to those of Brown, Black, and Horowitz, but their subjects were unable to match Japanese and Croatian words of the same meaning better than chance. They conclude that the latter finding is contrary to the Tsuru and Fries hypothesis of Gestalt organization of trace systems and related hypotheses of phonetic symbolism and physiognomic language such as that advanced by Brown, Black, and Horowitz. As in the Brown, Black, and Horowitz study, the concept population is indefinite and visual stimuli rather than auditory were used.

The hypothesis of a universal sound-meaning linkage based upon a primitive phonetic symbolism or physiognomic language is a nebulous one at best and does not seem to be

² As a check, the 42 words in the Brown, Black, and Horowitz list were presented to 28 psychologists with instructions to check those words which they believed named sense experiences. On only 16 of the words did 75 per cent or more of the subjects agree that the term named a sense experience.

susceptible to rigorous test in its general form. The results of the two studies cited appear, however, to constitute incontrovertible evidence that naive subjects can guess the English meanings of *certain* words from *some* foreign languages well above chance. The experiments reported below were designed to study aspects of the process of matching words of similar meanings from different languages. Specifically, we attempted to determine: (a) the effects on accuracy of guessing of mode of presentation (auditory, visual, or both), (b) whether the previously reported above-chance accuracy of guessing would occur when the word sample was a random one from a determinate population, and (c) the specific stimulus characteristics that lead to accurate guesses.

METHOD

Four major noncognate languages were used: Hebrew, Japanese, Chinese, and English. Choice of the first three was dictated largely by the availability of several skilled foreign-born translators for each of these languages.

The experimental population of words consisted of concepts of high frequency of occurrence. Ideally, a semantic frequency list of words for each language studied would form the basis for sample selection. However, since no such lists were available for the three foreign languages used, the closest approximation was to employ a list of the most frequently used concepts in four Indo-European languages: English, Spanish, French, and German (2). For the preliminary sample, those words were selected which were (a) principal parts of speech—either noun, pronoun, adjective, verb, or adverb, and (b) among the 500 most frequently used words in at least three out of the four languages. The 324 words obtained were typed and given to nine translators (three translators per foreign language). The translators were foreign-born, but were thoroughly familiar with the English language. They were instructed to work independently, to use the first person singular in translating verbs, to select the most commonly used expression when alternatives existed, and to use the colloquial rather than literary form. As a check on the appropriateness of the sample source, they were also asked to indicate those concepts which were uncommon in their own tongue.

On completion of each set of translations, the experimenters met with the three translators and compiled a composite list. Eliminated were those words marked as uncommon and those words upon which the re-translators could not easily reach agreement. The remaining terms were printed in romanized script by one of the translators and submitted by the experimenters to a fourth translator who back-translated them into English. In this way homonyms as well as other words which were not correctly back-translated were also eliminated. A final master list of words was drawn up

TABLE 1
VARIANCE TABLE FOR THE MATCHING OF
ENGLISH AND FOREIGN WORDS

Source	df	Sum of Squares	F	p
Individuals	117	1164.6		
Languages	2	167.9	7.46	<.001
Method of presentation	2	91.8	4.62	<.05
Languages $\times$ methods	4	72.9	1.62	>.10
Remainder	234	2636.6		
Total	359	4133.8		

by coordinating the separate lists for Japanese, Chinese, and Hebrew. From this alphabetized master list a sample of 50 words was taken by selecting every n th word.

A pilot study using an arbitrary list of 36 pairs of antonyms indicated that subjects did as well at guessing meanings when each English word was presented individually with either its foreign equivalent or opposite, as when the English and foreign antonyms were presented in pairs. In the interests of economy of both time and stimuli the simpler form was used in this experiment.

Three methods of presentation were used: auditory only (A), auditory and visual (AV), and visual only (V). Auditory presentation was by tape recording in which an announcer spoke the number of the word-pair and two native speakers the stimulus words. Approximately one second elapsed between presentation of the first and second words. An interval of five to seven seconds was allowed for the subject to make his decision. This interval had been found to be more than adequate in the pilot study mentioned above. For visual presentation, foreign words were typed in romanized form.

Each of the three methods of presentation was employed under two conditions. In one (E-F) the word pairs were composed of one English word and one foreign word: English-Japanese (E-J), English-Chinese (E-C), and English-Hebrew (E-H). In the other condition (F-F), one word came from one foreign language and the other word from a different foreign language: Chinese-Japanese (C-J), Chinese-Hebrew (C-H), and Hebrew-Japanese (H-J).

Six separate groups of subjects were used, corresponding to the six method-condition combinations. Each subject was confronted with three sets of 50 word-pairs. Order of presentation of the three sets was counterbalanced within each group of subjects.

Each set of 50 word-pairs consisted of 25, randomly selected, in which the two words meant the same thing and 25 in which they meant different things. Incorrect pairing was done by random assignment of the remaining words after the matched pairs had been selected. Separate randomizations were used for each language combination, although the same randomizations were used across methods of presentation.

Subjects were 269 undergraduates in an elementary psychology course. Eighteen were eliminated on the basis of knowledge of one or more of the foreign ex-

perimental languages, and eleven more were eliminated at random in order to equalize the size of all subject groups ($N = 40$ per group). All subjects were aware of the general purpose of the study and were informed that in only 25 word-pairs in each set of 50 did the two words mean the same thing. To increase involvement in the task, subjects were told that above-chance accuracy in guessing had already been demonstrated and that on completion of the task they could correct and score each other's papers. (Scoring was checked independently by the experimenters.)³

RESULTS

For each subject three scores were available, one for each of the three sets of 50 word-pairs. Scores were the number of correct guesses out of fifty. Two pseudo three-way analyses of variance were made of the data, one for the English-Foreign condition and one for the Foreign-Foreign condition. In each, individuals formed the rows, language combinations the columns, and methods of presentation (A, AV, and V) the blocks.

The results of the analysis for the English-Foreign language combinations are given in Table 1. Differences among languages were quite significant ($p < .001$) while differences among methods of presentation bordered on significance ($p < .05$). The differences among languages were determined by the relative ease with which the subjects guessed the equivalence of the English and Hebrew words, and among methods of presentation, by the superiority of the A-V and V methods to auditory presentation.

Table 2 presents the results of the analysis for the Foreign-Foreign language combinations. Again differences among languages were significant ($p < .001$) but those among methods of presentation were not. All differences among languages were significant. In this analysis there was a highly significant interaction effect with C-J declining in accu-

³ Two additional tables (a) a seven-page table giving the six sets of 50 word-pairs and the number of 40 subjects marking each word-pair as meaning the same thing under the three methods of presentation, and (b) a four-page table containing the list of artificial words and the frequency of selection of alternative response choices by 49 subjects have been deposited with the American Documentation Institute. Order Document No. 5093 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

racy with the addition of visual cues and C-H increasing in accuracy.

Tables 3 and 4 give the percentage of correct judgments for the several language combinations under the different methods of presentation. In the English-Foreign comparison, only E-H among the language pairs was guessed more accurately than chance, over all; among methods of presentation, both AV and V methods produced over-all above-chance guesses. In the Foreign-Foreign comparison, when methods of presentation are combined, the percentage of correct guesses for all languages is significantly different from chance, but in the C-H word-pairs the subjects would have done better if they had merely tossed a coin for their decisions. For methods of presentation, only the Audio form produced over-all above-chance results.

As a final analysis, the grand totals of correct judgments under the two conditions (E-F and F-F) were compared. The number of correct judgments made in the Foreign-Foreign comparisons was not significantly greater than that made in the English-Foreign.

These results leave the answer to the first question raised—i.e. the effect of method of presentation on accuracy of guessing—somewhat uncertain. In the E-F comparison visual cues tend to raise the number of correct guesses; in the F-F comparison the effect of adding visual cues is variable depending upon the language pairs.

The second question, i.e. can subjects pair words from two different languages with above-chance accuracy when the sample of words is a random one from a specified population, can be answered affirmatively—for some language pairs. There was a wide range in guessing accuracy for the six language combinations with Chinese-Japanese word-pairs being matched most accurately and Chinese-Hebrew word-pairs the least accurately. In general, the significance levels for all comparisons are much lower than previously reported.

The third question, i.e., what are the cues that lead to correct guessing, required an additional analysis of the data. It is apparent that above-chance accuracy is determined by the fact that, on the average, the agreement among the subjects on the sameness or differentness of the word-pairs tended to correspond to objective sameness or differentness in word meanings. Consequently, it was possible to

TABLE 2
VARIANCE TABLE FOR THE MATCHING OF
WORDS IN TWO FOREIGN LANGUAGES

Source	df	Sum of Squares	F	p
Individuals	117	1271.3		
Languages	2	691.3	31.83	<.001
Method of presentation	2	6.7	.31	>.10
Languages X methods	4	145.8	4.02	<.01
Remainder	234	2122.3		
Total	359	4237.4		

TABLE 3
PERCENTAGE OF CORRECT JUDGMENTS OF
THE EQUIVALENCE OF MEANINGS OF
ENGLISH AND FOREIGN WORDS

Languages	Method of Presentation			Total
	Audi-tory	Audio-visual	Visual	
English-Japanese	47.2*	52.2	51.4	50.3
English-Hebrew	52.6	53.3*	53.0*	53.0*
English-Chinese	49.0	49.9	50.8	49.9
Total	49.6	51.8*	51.7*	51.0*

* $p < .01$.

TABLE 4
PERCENTAGE OF CORRECT JUDGMENTS OF THE
EQUIVALENCE OF MEANING OF WORDS
FROM TWO FOREIGN LANGUAGES

Languages	Method of Presentation			Total
	Audi-tory	Audio-visual	Visual	
Chinese-Japanese	57.0*	53.8*	53.7*	54.8*
Chinese-Hebrew	46.8*	47.7	49.9	48.1*
Hebrew-Japanese	52.6	53.3*	50.9	52.3*
Total	52.1*	51.6	51.5	51.7*

* $p < .01$.

approach the problem of the cues that lead to accurate guessing by studying the factors that led to agreement among the subjects, whether or not such agreement was in fact correct on any particular word-pair.

There are undoubtedly many factors that lead to intersubject agreement. However, the most promising set of determinants for analysis lies in the objective stimulus characteristics of the word-pairs. To determine these cues, the individual responses of each group of 40 subjects for each language combination were tabulated and those word-pairs isolated on

which more than 70 per cent of the subjects agreed in marking them as "same" or "different." Such high percentages of agreement were quite common, averaging more than 20 out of each set of 50 word-pairs. The list of word-pairs thus isolated was inspected for cues that might have led to the high degree of agreement observed. A number of possible factors was apparent; the five that seemed to predominate were:

1. Length—words containing the same number of letters and/or syllables tended to be marked "same" and those of differing lengths "different."

2. Vowels—words containing exclusively or predominantly vowels of the same group (*a-o-u* or *i-e*) tended to be marked "same" whereas when the vowels in the two words were from different groups, they were judged "different."

3. Consonants—words containing exclusively the same consonant types (sibilants, explosives, etc.) were marked "same." Differing consonant types in the two words elicited the "different" response.

4. Hyphenation or spacing—where perceptible spacing occurred in both words they tended to be marked "same."

5. Connotation—if the two words or parts thereof suggested a single English concept, the two words tended to be marked "same."

These factors were not all equally frequent nor present to equal extent in the differing language pairs. Connotation, for example, appeared to be much stronger in the English-Foreign comparisons than in the Foreign-Foreign. Similarly, certain of the factors appeared to have considerable strength in eliciting a "same" response but did not affect the "different" judgment as markedly. The converse was also true.

In order to determine the relative potency of the postulated factors, the original six lists of 50 word-pairs were checked for the occurrence or nonoccurrence of the factors in each word-pair. Where a factor was present in a given word-pair and the largest frequency of response was in the direction expected, a plus tally was made for that factor. If the largest frequency of response was in the direction opposite that expected, a minus tally was made. A comparison of the proportion of plus tallies to total tallies for each factor gave the

following potency ranking: connotation, vowels, length, spacing, and consonants.

Since the analysis described above was highly subjective, a second experiment was conducted to test the postulated effects of three of the factors described above (connotation, length, and vowels.) To control the occurrence of the factors, artificial "languages" were used, the words of which were constructed from nonsense syllables. Presentation of the words was visual only with each item consisting of one stimulus "word" paired with two response "words." The subject's task was to select that one of the alternatives which meant the same thing as the stimulus "word."⁴

The alternative choices in each item were constructed so as to match or to differ from the stimulus word in one or more of the three factors under consideration. Table 5 indicates the factors in which the two response words matched the initial word, the number of items of each type, an example of each type of item, and the percentage of 49 subjects who selected the predicted alternatives in each type of item. The last column indicates the rank order of potency predicted on the basis of the original derivation of the factors. Two rank orders of potency were predicted, one where one factor at a time was investigated while the others were held constant, and one for the factors pitted against each other.

In the construction of the words, nonsense syllables of low association value were used where connotation was to be minimal. For the words with high connotation value, nonsense syllables of high association value were used in part as well as some nonsense words constructed from the Kent-Rosanoff list of high associative stimulus words. In words in which vowels were matched, each vowel within one group (*a, o, u, or i, e*) was matched with itself and with its fellow members an equal number of times. In the words in which vowels did not match, *i* and *e* were paired equally frequently with *a, o,* and *u*. In the response words where length was to be different from the initial word, the response word was equally frequently longer and shorter. Differences in length were varied from two to five letters.

The data in Table 5 show one reversal from the predicted rank order of potency. The factor

⁴ See Footnote 3.

TABLE 5
FACTORS ON WHICH RESPONSE "WORDS" WERE MATCHED WITH STIMULUS "WORDS" AND PERCENTAGE OF 49
SS SELECTING THE FIRST ALTERNATIVE
(Artificial "words" formed from nonsense syllables)

Response Words		No. of Items	Example	Percentage of Subjects Selecting A	Predicted Rank Order of Potency
A	B				
Vowels Length Connotation	Vowels Length	9	famm hoam noax	80	1
Vowels Length	Length	30	jiaq yil yom	66	2
Vowels Length	Vowels	9	wog zos zotyom	71	3
Vowels Length Connotation	None	9	hur wom wibpid	85	1
Vowels Length	None	9	yod zoj xezhief	82	2
Vowels Connotation	Length	8	wat lakcull jid	65	3
Length Connotation	Vowels	9	hevv wayt meijzieb	66	4
Length	Vowels	8	wez gebeviff xag	57	5
Vowels Length	Connotation	9	darq neitz naox	56	6

of length turned out to be more effective in determining the subjects' decisions than was the factor of vowel similarity. All the percentages of agreement obtained were significantly greater than chance with the exception of that obtained in pitting C against L plus V. These results support the conclusions drawn from the empirical item analysis of the natural languages. Similarity in length of words, vowel groups, and connotation significantly influence the decisions of subjects as to similarity of meaning.

DISCUSSION

Certain discrepant results appear among the two studies cited and that reported here. Brown, Black, and Horowitz found that their subjects were able to guess the English meanings of Chinese words significantly above

chance, whereas we did not. Maltzman, Morrisett, and Brooks report that their subjects were able to guess the English equivalents of Japanese words in a similarly successful fashion; we found that this was not the case for our subjects. There appear to be three possible sources for this disagreement: (a) method of presentation (audio, audiovisual, or visual), (b) design of stimulus presentation—(simple paired comparison, single stimulus—double response choice and double stimulus—double response choice), and (c) the word sample used. The first source appears to be ruled out, for where we used the method of presentation employed by the other investigators, negative results were still obtained (Table 3). On the basis of the pilot study briefly mentioned, the difference between designs of stimulus presentation likewise did

not appear to be crucial. The third alternative thus remains as the most likely one although an interaction effect between (b) and (c) cannot be ruled out. According to this interpretation, it is possible to match words for meaning between English and certain foreign languages with better than chance success, without knowledge of the foreign equivalents. The process is facilitated by marked contrasts of meaning such as occur in antonym pairs or contradictory word pairs. Where a random sample of words from a general concept population is used and where meaning contrasts are not as great, correct guessing becomes much more difficult although it may still occur with better than chance success in some languages, e.g., English-Hebrew.

Maltzman, Morrisett, and Brooks also found that their subjects were unable to match Japanese-Croatian words of the same meaning whereas in the present study the subjects were able to pair Chinese-Japanese words and Hebrew-Japanese words significantly better than chance. However, our subjects were significantly poorer than chance on Chinese-Hebrew words. The conclusion of Maltzman *et al.*, that their results tend to disprove the general hypothesis of universal phonetic symbolism, would seem to be correct, but the positive findings observed raise certain interesting alternatives. The observation that noncognate languages may employ similar patterns of sounds (and length) to designate similar meanings—at least among frequently used words—offers an intriguing way of grouping languages. However, it should be emphasized that the restricted nature of the population of concepts used in this study does not permit conclusions about the degree of similarity between *total* word populations of any two of the languages used.

SUMMARY

Six groups of 40 English-speaking subjects were presented with the task of judging the equivalence of meaning of three sets of 50 pairs of words presented auditorily, audiovisually, or visually. The words were a random

sample of concepts of high frequency of usage and the three sets were composed of pairs of words in English and Chinese, English and Japanese, and English and Hebrew, or Chinese and Japanese, Japanese and Hebrew, and Chinese and Hebrew.

The subjects were able to guess the English equivalents of Hebrew words and to pair Japanese and Chinese, and Japanese and Hebrew words at better than chance levels. In pairing Chinese and Hebrew words the number of correct guesses was significantly below chance. In the English-Foreign comparisons, visual presentation of the romanized versions of the foreign words tended to increase the percentage of correct guesses; but not in the Foreign-Foreign comparison. A significant interaction effect between languages and method of presentation was noted in the latter situation.

An item analysis of the subjects' responses indicated that similarity of word lengths, vowels, consonant types, spacing of compound words, and connotation markedly influenced the agreement among the subjects as to the sameness of meaning of word pairs. A second experiment indicated that agreement among subjects as to the meaning of synthetic words could be systematically varied by manipulation of the cues of length, vowels, and connotation.

The results of the study are interpreted as contrary to the hypothesis of a universal phonetic symbolism.

REFERENCES

1. BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages. *J. abnorm. soc. Psychol.*, 1955, **50**, 388-393.
2. EATON, H. S. *Semantic frequency list for English, French, German, and Spanish*. Chicago: Univer. of Chicago Press, 1940.
3. MALTZMAN, I., MORRISSETT, L., & BROOKS, L. O. An investigation of phonetic symbolism. *J. abnorm. soc. Psychol.*, 1956, **53**, 249-251.
4. OSGOOD, C. E. & SEBEOK, T. H. (Eds.) *Psycholinguistics. A survey of theory and research problems*. *J. abnorm. soc. Psychol.*, 1954, **49** (4, Pt. 2—Suppl.).
5. TSURU, S. & FRIES, H. S. A problem in meaning. *J. gen. Psychol.*, 1933, **8**, 281-284.

Received May 31, 1956.

PERFORMANCE ON THE BENDER UNDER HYPNOTIC AGE REGRESSION

HAROLD B. CRASILNECK AND CARMEN MILLER MICHAEL

Southwestern Medical School, University of Texas

CONSIDERABLE debate may be found in psychological literature concerning the phenomenon called "hypnotic age regression." Some reports (1, 4, 5, 6, 7, 8, 11, 12, 14) have indicated that a subject can be taken back hypnotically to a given age and that his behavior then is exactly as it was at that age—that there is a genuine regression involving the total organism. Other studies (2, 3, 9, 10, 13, 15, 16, 18) suggest that hypnotically regressed subjects attempt to play an assigned role and to cooperate with the hypnotist's suggestions, but do not actually, organismically reproduce the earlier behavior. The present study addresses itself to this debate.

In the various experiments that have been designed to test the "role-playing" versus the "true regression" hypothesis, subjects have been given a variety of physiological and psychological tests. In such studies, particularly when psychological tests are used as a criterion of regression, the subject should not be aware of how an individual at the assigned age should behave. The writers believe that the Bender Motor Gestalt Test is one which allows for less of such awareness than has usually been the case in tests used in previous studies.

In the present experiment, the problem was to determine the absolute and relative levels of maturity of behavior of subjects under four kinds of conditions, and to test especially the hypothesis that under hypnotic age regression, subjects behave exactly as they did at a given, suggested age. For the reason previously mentioned, the Bender Motor Gestalt Test was selected as the instrument to be used in judging maturational level.

METHOD

Subjects

The Ss were ten white female student nurses who had volunteered their services for this experiment. They ranged in age from 19 to 22 years with a mean of 20. Although the Ss knew that the experiment "had something to do with hypnosis," they were naive concerning the specific problem involved. In each case S was repeatedly hypnotized over a period of time until it was definitely ascertained that she was a somnam-

bule, capable of manifesting those signs described by Wolberg (17). Prior to the experimental procedure either the Wechsler-Bellevue Intelligence Scale (Form I) or the Wechsler Adult Intelligence Scale was administered by one of the staff psychologists not associated with the hypnotic sessions.¹

Procedure

The Bender was given to each S individually under four conditions in the following sequence: (a) awake, (b) awake with instructions to pretend to be four years old, (c) hypnotized and instructed to pretend to be four, and (d) hypnotically regressed to four.

Condition 1. After entering the experimental room, S was seated at a desk, handed a pencil and a sheet of paper, and given the following instructions: "I have some cards like this and I want you to draw some things for me. Now make one like this and when you are through, I'll give you the next one." After completing the Bender, S was asked to put her name on the paper.

Condition 2. The S was then told, "Now I want you to pretend that you are four years old. I want you to do anything that is asked of you just like you would have done it when you were four years old." The S was again handed a pencil and paper, given the Bender, and asked to put her name on the paper.

Condition 3. Following this, S was hypnotized by the visual fixation technique into a somnambulistic state. She was then told that, "In a moment you will be able to open your eyes; however you will still remain deeply sound asleep. Upon opening your eyes, I want you to pretend that you are four years old. I want you to do anything that is asked of you just like you would have done it when you were four years old." Upon opening her eyes, S was again given the Bender and asked to put her name on the paper.

Condition 4. The S was then told to close her eyes and was hypnotically regressed to four years of age. During the regression, she was asked questions concerning her age, background, and professional knowledge. After it was reasonably established that S was manifesting the usual, expected signs of regression to four years of age, it was suggested that she would open her eyes, but remain sleeping, and that she would do what was asked of her. The Bender was then administered and S was told to put her name on the paper. The S was then hypnotically returned to her present age with amnesia for the entire episode.

Judgments of Maturation Level

Three clinical psychologists who were completely naive as to the nature of the experiment served as

¹ The authors wish to thank Dr. Arthur Lee Foster and John Gladfelter.

judges.² The 40 Bender protocols were disguised with regard to identity and arranged in a random order. They were given to the judges individually with the following instructions:

Here are 40 Bender Gestalts. They are numbered from 1-40. You will notice that each is on a separate sheet of paper, except for #28 and #34, each of which is on two sheets.

Each Bender has a name and a date and some other information on it. All of this has been covered over by some heavy blue paper. It is essential that you leave the blue paper in place and not look at the information underneath.

You are requested to make a judgment concerning each of these Benders. Using your past experience and the help of Plate 62 in Bender's Monograph, estimate the maturation level in the visual motor gestalt function of each Bender. Each judgment should be one of the following: $3\frac{1}{2}$ (under 4 years), 4, $4\frac{1}{2}$, 5, $5\frac{1}{2}$, 6, $6\frac{1}{2}$, 7, $7\frac{1}{2}$, 8, $8\frac{1}{2}$, 9, $9\frac{1}{2}$, 10, $10\frac{1}{2}$, 11, or $11\frac{1}{2}$ (over 11 years).

Please use the attached Judgment Sheet to record your judgments. Your comments are not being solicited, but if you have any, you may make them in the space provided.

The protocols were then rearranged into ten sets. Each set consisted of the four records from each *S* in a random order. These materials were again given to the judges individually with the following instructions:

Here are 10 sets of Bender Gestalts. There are four Benders in each set. Each set is from a different subject, but all four Benders in each set are from the same subject. Each Bender is on a separate sheet of paper, except for #28 and #34, each of which is on two sheets.

Each Bender has a name and a date and some other information on it. All of this has been covered over by some heavy blue paper. It is essential that you leave the blue paper in place and not look at the information underneath.

The Benders are numbered randomly from 1-40. Each Bender also has a name (initials). Of course, in each set of four all the Benders will have the same name.

You are requested to rank the four Benders in each set according to your estimate of the relative maturational level in the visual motor gestalt function. Please rank as "a" the one which you think is most mature, "b" the next most mature, "c" the next, and "d" the one which you think is most immature. Use your past experience and the help of Plate 62 in Bender's Monograph.

Please use the attached Judgment Sheet to record your judgments. Your comments are not being solicited, but if you have any, you may make them in the space provided.

The reliability of the judges' absolute ratings of the maturational level of the 40 Benders was determined by calculating the intercorrelations of judgments by the method of rank-difference. The correlation coefficients were .90, .84, and .89. Since the reliability of the

judges could be regarded as high, their absolute judgments were averaged for each case and their composite ratings used for the analysis of the data on maturational level.

RESULTS

The mean maturational level of each of the four conditions under which each *S* took the Bender was determined. The means for Conditions 1 through 4 were, respectively, as follows: 11.2 years, 9.9 years, 7.8 years, and 7.3 years.

The significance of the difference between any two of the above means was determined by computing the significance of the difference between correlated means. All except two of the differences were significant at the .01 level of confidence. The difference between Mean 1 and Mean 2 was significant at the .02 level. The difference between Mean 3 and Mean 4 was not significant. When the significance of the difference between each of the four means and the hypothetical mean of four years was assessed, all differences were significant at the .01 level.

On the basis of the *Ss'* Wechsler-Bellevue and WAIS IQ scores, estimates were made of their probable mental ages when they were chronologically four years old. The Wechsler scores were translated into Standard scores and then converted into comparable Stanford-Binet IQs for four-year olds. The mean estimated mental age thus obtained was 4.6 years. The difference between each of the means for Conditions 1 through 4 and this index of mental age (4.6 years) was determined by computing the significance of the difference between correlated means. Each of the four differences was significant at the .01 level of confidence.

These results with absolute ratings of maturational level led to the following question: If the judges make *relative* ratings, i.e., if they rank each of the four Benders from the same *S* in order of maturity, will the difference between the rankings under the different conditions, especially between Conditions 3 and 4, be significant? The judges' rankings were averaged for each case, and the significance of differences was determined by the method of paired replicates. The findings were that the difference between rankings under any two conditions was significant at the .01 level *except* the difference between the third and fourth conditions, which was not significant. The means of the judges' rankings from most to

² The authors wish to express gratitude to Mrs. Genette Burruss, Dr. Rosaline Goldman, and Dr. Harold Winer.

least mature under the four conditions were 1.1, 2.0, 3.3, and 3.6, respectively.³

DISCUSSION

The foregoing results indicate that when *Ss* are asked to perform at a given lower age level, they comply to a degree but do not seem to reach the actual age level suggested. They behave more nearly like the given age when the suggestion is made when *S* is under hypnosis than when in a waking state. When in a state of hypnotic age regression, some of the *Ss* tended to behave more nearly like the suggested age than they did while hypnotized but merely instructed to pretend they were the suggested age. However, the difference in performance under these two conditions was not found to be consistent or statistically significant. It appears, then, that the results do not support the idea that hypnotically regressed *Ss* behave exactly as they did at the suggested age.

Since all *Ss* were given the Bender under four conditions in the same sequence, it is possible that they may have learned by practice and that therefore their performance level was not as low as it would have been had they not had the practice. Even if this is true, however, it would still support the inference that a total regression does not take place.

One may, of course, question the authenticity of the hypnotic technique and the criteria which were used as evidence that the *Ss* were actually in a somnambulistic state and capable of demonstrating regression as well as could be expected. Each of the *Ss* manifested those signs of somnambulism and regression listed by Wolberg (17), and during the process of age regression, all *Ss* manifested personality changes of a regressive nature. For example, handwriting, pencil grip, vocabulary, and pronunciation of words became more and more childish. When hypnotically regressed to the age of four, one *S* began to cry when she was urged to put her name on the paper. When questioned she said she was crying because she did not know how to write her name. Four *Ss*, who did not ordinarily bite their nails, began to do so when hypnotically regressed to an early age. Still other signs, usually accepted

as evidence of age regression, were obtained after the actual experimental data were collected. For example, while still in a state of hypnotic age regression, the *Ss* were given the suggestion that they would return to the age of one year. They then no longer answered when asked questions. Some *Ss* were further regressed to four months of age at which time the sucking reflex was elicited.

Although the Benders obtained while the *Ss* were hypnotically regressed (to the age of four years) were reliably rated by the judges to reflect an average maturational level of 7.3 years, it should not be concluded that their performance was typical for the age of seven or that an organismic regression down to seven years actually occurred. All of the judges spontaneously commented that it was difficult for them to assign an absolute maturational level to the Benders because there was such marked intratest variability on most of them. One of the judges rated each gestalt separately and then took the median rating as the best approximation of over-all functioning level. Another judge reported that, "Some of the subjects were children; some were adults; some were psychotic; and some were organic." The judges also indicated that it was much easier to make relative judgments than absolute judgments.

Although the results indicate that hypnotically regressed *Ss* do not actually behave totally and exactly as they did at the suggested age, one may legitimately speculate about what might account for the behavioral changes which do occur. The *Ss* appear to be attempting to follow the hypnotist's suggestions as exactly as they can, and hypnotic techniques seem to increase their desire or to facilitate their attempts to follow suggestions, or, as some investigators have called it, to play an assigned role. But in addition to the probable role-playing aspects of behavior under hypnosis, the possibility still exists that fragments of earlier behavior may be reproduced during hypnotic regression. There may occur, for example, a reinstatement of earlier reflex or conditioned reflex behavior even though a total regression does not occur.

There are two methodological issues, however, which limit the interpretation to be placed on the present data. First, maturational level was defined in terms of judges' ratings of

³ The authors are indebted to Dr. Alvin J. North for his help in setting up the statistical design of this research.

Bender protocols. Like all ratings, these estimates may have been subject to the error of central tendency, the probability that judges will use the middle of the range rather more than the extremes. Since the range of possible mental age ratings ran only from four to eleven, this error of central tendency, if it actually occurred, could conceivably account for the significant differences between the obtained values and the hypothetical base of 4.6 years. Only the absolute judgments, of course, would be affected, the relative comparisons being free from any influence from this source.

Second, what has been said here assumes that ratings by qualified judges of Bender protocols constitute a legitimate and useful operational definition of maturational levels. Other defining operations might well lead to rather different findings. Nevertheless, the present procedure seems to be strongly defensible on at least two counts: (a) the high reliability of the judges indicates at least a great deal of commonality in clinical opinion about maturational level as revealed by the Bender, and (b) Ss are probably less able to imitate, either consciously or unconsciously, their performance at presumably lower age levels on the Bender than is the case with other indices of maturational development. Consequently, it would appear that the method used was a sound one in the light of the hypothesis tested.

SUMMARY

In this study, the problem was to determine the absolute and relative levels of maturity of behavior of Ss under four conditions, and especially to find whether hypnotically regressed Ss behave exactly as they did at a suggested age. The Bender Motor Gestalt Test was selected as the instrument to be used in judging maturational level. Ten Ss were administered an intelligence scale. These Ss then took the Bender under the following conditions: awake, awake and pretending to be four years old, hypnotized and pretending to be four, and hypnotically regressed to four. Independent ratings by three judges revealed,

among other things, that when Ss are asked to perform at a given lower age level, they comply to a degree, but do not seem to reach the actual age level suggested.

REFERENCES

1. BERGMAN, M. S., GRAHAM, H., & LEAVITT, H. C. Rorschach exploration of consecutive hypnotic chronological age level regressions. *Psychosom. Med.*, 1947, **9**, 20-28.
2. FARBEROW, N. L., & SARBIN, T. R. A clinical study of role and self in hypnotic age regression. *Amer. Psychologist*, 1950, **5**, 305. (Abstract)
3. KEIR, G. An experiment in mental testing under hypnosis. *J. ment. Sci.*, 1945, **91**, 346-352.
4. KLINE, M. V. Hypnotic age regression and intelligence. *J. genet. Psychol.*, 1950, **77**, 129-132.
5. KLINE, M. V. Hypnosis and age progression: a case report. *J. genet. Psychol.*, 1951, **78**, 195-206.
6. KLINE, M. V. Childhood fears in relation to hypnotic age regression: a case report. *J. genet. Psychol.*, 1953, **82**, 137-142.
7. KLINE, M. V. Hypnotic retrogression: a neuropsychological theory of age regression and progression. *J. clin. exp. Hypnosis*, 1953, **1**, 21-28.
8. LEEDS, M. An hypnotic regression series. *Persona*, 1949, **1**, (1), 13-16.
9. MCCRANIE, E. J., CRASILNECK, H. B., & TETER, H. P. The electro-encephalogram in hypnotic age regression. *Psychiat. Quart.*, 1955, **29**, 85-88.
10. ORNE, M. T. The mechanisms of hypnotic age regression. *J. abnorm. soc. Psychol.*, 1951, **46**, 213-225.
11. PLATONOW, K. I. On the objective proof of the experimental personality age regression. *J. gen. Psychol.*, 1933, **9**, 190-209.
12. SARBIN, T. R. Mental age changes in hypnotic age regression. *Amer. Psychologist*, 1946, **1**, 456-457. (Abstract)
13. SARBIN, T. R. Mental age changes in experimental regression. *J. Pers.*, 1950, **19**, 221-228.
14. SPIEGEL, H., SHOR, J., & FISHMAN, S. An hypnotic ablation technique for the study of personality development. *Psychosom. Med.*, 1945, **7**, 273-278.
15. TAYLOR, A. The differentiation between simulated and true hypnotic regression by figure drawings. Unpublished master's thesis, City Coll. of New York, 1950.
16. TRUE, R. M. Experimental control in hypnotic age regression states. *Science*, 1949, **110**, 583-584.
17. WOLBERG, L. R. *Medical hypnosis*. Vol. 1. New York: Grune & Stratton, 1948.
18. YOUNG, P. C. Hypnotic regression—fact or artifact? *J. abnorm. soc. Psychol.*, 1940, **35**, 273-278.

Received May 28, 1956.

DECISION PROCESSES IN COMMUNICATION NETS¹

MARVIN E. SHAW, GERARD H. ROTHSCILD, AND JOHN F. STRICKLAND

The Johns Hopkins University

THERE is ample experimental evidence that the communication structure of a group influences the performance and satisfaction of the group when the task is to solve problems having logically imperative solutions (2, 8, 10, 12, 14, 15). However, there seems to be no evidence that the same kinds of effects obtain when the problem to be solved does not have a generally agreed upon solution, e.g., a problem requiring agreement among the group members about the proper way to deal with an issue involving interpersonal relations. The first experiment reported in this article attempted to determine the effects of certain communication nets upon group performance and satisfaction when the task is to solve such a "human relations" problem. Experiments II and III were attempts to answer certain questions raised by the results of Experiment I.

EXPERIMENT I

Method. The apparatus used in this experiment is the same as that described in earlier reports (14, 15). It consists of four cubicles, each of which has a worktable, work materials, and a switch which controls a light and timer located at *E*'s position. The cubicles are interconnected by means of slots through the walls separating them. Group members can communicate with each other by writing messages on cards and passing these cards through the slots. Various communication nets can be arranged by closing the necessary slots. The communication nets investigated in this study are shown in Fig. 1.

The *Ss* for this experiment were students (all male) at The Johns Hopkins University. All were highly experienced, having served previously as *Ss* in this kind of experimental situation for a total of 12 sessions during which they solved a total of 24 problems involving simple arithmetical manipulations. For purposes of counterbalancing, the experimental design required that eight groups of four *Ss* be assigned to each of the three communication nets. One group which had been assigned to the "slash" failed to report for the experimental session, and since no other experienced *Ss* were available the data reported here are for eight groups in the "star," eight in the "comcon," and seven in the

"slash." It appears highly unlikely that failure to fill this one cell in the design resulted in any serious bias in the final results.

Two human relations problems were used. One was a modified version of Deutsch's "The World War II Veteran" problem (3). As presented to the *Ss* in this experiment, the central figure in the story was an executive who had had an affair with his secretary. The secretary had subsequently resigned to get married. The executive's wife is intelligent but suffers from a traumatic childhood experience which made her especially sensitive to this kind of unfaithfulness. The question: What should the executive do, assuming that for his own peace of mind he cannot just forget the matter?

The other problem was essentially the same as that used by Festinger and Hutte (6). The central figure in this story is Henry, the son of a physician, who has a friend, Jim, who is incurably ill. Both are in love with the same girl. Just as Henry has decided to ask the girl to marry him, she announces her engagement to Jim. The girl does not know that Jim is incurably ill, although Henry does. The question: What should Henry do and say?

Except for the communication net and the order in which the problems were presented, the procedure was the same for all groups. The *Ss* were seated in their cubicles and handed a mimeographed set of instructions. The *E* then read the instructions while the *Ss* followed along on the printed page. The instructions informed *Ss*: (a) of the nature of the communication net (which was, with minor exceptions, the same as during previous sessions), (b) that they were to solve problems having no correct or incorrect answer, but which merely called for a group decision, (c) that a unanimous decision was desirable but that they could give a minority report if they were convinced that unanimity could not be attained, and (d) that they would be given 20 minutes to discuss the problem, but could signal *E* by throwing their switches if agreement was reached before the end of this period. The lack of emphasis upon time is in contrast to instructions given in previous investigations. Problems were presented in counterbalanced order.

After both problems had been considered and decisions rendered, *Ss* were required to indicate on an 11-point scale how well they enjoyed their task in the group and how they would rate the performance of their group.

Results. Of chief interest were the differences due to the communication net and to positions within nets. In addition, data were evaluated

¹ This experiment was done under Contract N5-ori-166, Task Order 1, between the Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-198, Project Designation No. NR 145-089, under that Contract.

² The second problem appeared to be much more interesting to the *Ss* in this experiment, perhaps because most were unmarried and hence the situation tapped a more relevant frame of reference.

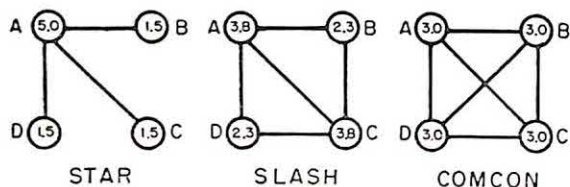


FIG. 1. THE EXPERIMENTAL NETS

Circles represent persons and the lines represent two-way communication channels between persons. Numbers within circles are the Independence scores (13) for the positions within the net.

TABLE 1
MEAN NUMBER OF COMMUNICATION UNITS
TRANSMITTED BY Ss IN DIFFERENT POSITIONS

Nets	Positions				<i>p</i>
	A	B	C	D	
Star	50.8	22.8	18.1	21.8	< .01 ns ns
Slash	41.6	29.6	38.3	29.0	
Comcon	31.5	33.3	32.8	32.6	

for differences due to trials (first vs. second problem) and to problems. In agreement with the results of previous studies, trials differed significantly with respect to both time and message units required to reach a decision ($p < .01$). This finding is a general one and will not be referred to again. Differences due to problems were not significant in any of the analyses performed.

There were marked differences among the groups in different communication nets in mean times required to reach a decision (16.62 min. in the star, 14.00 min. in the slash, and 9.71 min. in the comcon). The F test revealed that these differences are statistically reliable ($p < .01$). This finding is in agreement with the results obtained when the problems to be solved were of the arithmetical type, but it might be noted that the magnitudes of the differences are somewhat greater than those typically reported in earlier investigations.

There were no significant differences among positions within nets with respect to time scores. In fact, under the conditions of this experiment it would have been very surprising if such differences had been found.

Content of messages was analyzed into units by the method developed in earlier studies (12, 14, 15) according to which a message unit is defined as any simple sentence or any meaningful part of a complex or compound sentence. There were no significant differences among

nets in number of message units required to reach a decision (means were 28.4, 34.6, and 32.6 for the star, the slash, and the comcon, respectively). This does not agree with the findings of earlier studies, but this failure seems to be due almost entirely to the great differences in time required to reach a decision. If message units per unit time are considered, nets are ordered in the same way as in other experiments. (Mean message units per min. were 3.60, 4.95, and 7.19 for the star, the slash, and the comcon, respectively. Differences are statistically reliable [$p < .001$].)

Mean number of message units transmitted by Ss in the various positions within nets are shown in Table 1. Only the differences in the star are significant, but the differences in the slash are in the same direction as that found in earlier studies. That is, Ss in the more centralized positions sent more messages than did Ss in the more peripheral positions. It is also interesting that the number of messages directed toward deviators (i.e., those who disagreed with the majority opinion) was considerably greater than the number directed toward nondeviators in equivalent positions (mean number message units received by deviators: 11.9, 19.3, and 22.7 in the star, the slash, and the comcon, respectively; by non-deviators: 9.2, 14.3, and 16.9 in the star, the slash, and the comcon, respectively). Differences between deviators and nondeviators were significant at well beyond the .01 level of confidence in each of the three nets. This finding agrees with results obtained under very different conditions by Festinger and Thibaut (5) and by Schacter (11).

There were no significant differences among nets or among positions within nets either in ratings of satisfaction or in ratings of performance. However, it is interesting that the direction of the differences in ratings of satisfaction were opposite to that found in previous investigations (mean ratings of satisfaction: 9.09, 8.86, and 8.63 for the star, the slash, and the comcon, respectively; in the star, mean ratings of satisfaction were 8.3 for Position A and 9.4 for Positions B, C, and D, where higher numbers indicate greater satisfaction). A possible explanation for this result is given in a later section.

The frequency with which individuals disagreed with the majority opinion was very small in all three nets; hence, application of

statistical tests was not justified. However, it is interesting to compare the percentage of disagreements occurring in the three nets. The percentage of total opinions (decisions) which disagreed with the majority of the particular group was 18.75 in the star, 16.07 in the slash, and 9.38 in the comcon. Thus, despite the fact that Ss in the star and the slash spent more time trying to reach an agreement, the degree of unanimity actually attained was less than in the comcon.

Differences in disagreements by Ss in the various positions within nets are shown in Table 2. It can be seen that frequency of disagreement correlates rather well with the centrality of positions. In the star the most central position never disagreed with the majority opinion. Likewise the more central positions in the slash showed markedly fewer disagreements than did the more peripheral positions.³

Discussion. Taken together, the results given above present a reasonably consistent picture of the group process. The time required to reach a decision was longest in the star and shortest in the comcon, in agreement with earlier studies in which arithmetical problems were used. Number of message units required to reach a decision in the three nets did not differ significantly, but units transmitted per minute did. In other words, groups in the comcon reached a decision more quickly than did groups in the other two nets but did so by transmitting messages at a more rapid rate.

It has been suggested that the time differences are due entirely to the fact that in the star and in the slash the burden of writing falls mainly on one or two people; i.e., that differences are due entirely to the effects of the net upon ease of intragroup communication. Fortunately, this hypothesis can be checked rather easily. If it is true, then the rate of message transmission should be essentially the same for the S sending the greatest number of messages in each net. This is not the case; the star is significantly slower than either the slash or the comcon ($F = 6.85, p < .01$). Also, it would be expected that if we selected the one S in each group who sent the greatest number of messages for that group, and then compared nets with respect to number of

³ The slight aberration in the comcon is due largely to the fact that one group split two-two; in this case, two positions were arbitrarily selected as being deviant.

TABLE 2
PERCENTAGE OF DECISIONS DISAGREED WITH BY
Ss IN DIFFERENT POSITIONS

Nets	Positions			
	A	B	C	D
Star	0	31.3	25.0	18.8
Slash	7.1	28.6	7.1	21.4
Comcon	6.3	12.5	6.3	12.5

messages transmitted by these Ss alone, nets would differ significantly. This is not found ($F = 1.31; F_{.05} = 3.49$). Finally, if this hypothesis were true one would expect that time, measured from the "go" signal until the last S had thrown his switch, would correlate significantly with the greatest number of messages sent by any one S in that group. Again, the result is negative ($r = .097; r_{.05} = .396$). The results of these analyses make it clear that the interrelationships of nets, decision times, and messages are not simple ones.

Another finding that requires some discussion is the relationship between net and frequency of disagreement with the majority opinion. There are two possible explanations. One is that the Ss in the more central positions were using their position to bring other members around to their own viewpoint. A second possibility is that Ss in the more central positions who held deviant opinions were subject to more direct pressure than were those in peripheral positions; this added pressure caused them to change more quickly and more frequently than peripheral deviators. The latter interpretation also explains the fact that ratings of satisfaction by Ss in different positions were not significantly different. The peripheral positions may have enjoyed being able to maintain a deviant opinion without being subjected to too much group pressure; on the other hand, Ss in the more central positions may have had their satisfaction dampened by being vulnerable to group pressure.⁴

The results of Experiment I do not permit a choice between these two alternatives. Ex-

⁴ Another factor in the determination of ratings of satisfaction is saturation (8), i.e., the degree to which the channel capacity is reached or exceeded. When the communication demands upon persons in central positions become too great, their satisfaction with the position is reduced. The saturation point was reached in the star and to some extent in the slash.

periment II was designed to shed further light upon this question.

EXPERIMENT II

Purpose. This second experiment was designed to examine further the effects of the position within a communication net upon the ability of the *S* in the position to maintain an opinion in the face of opposition. Specifically, we wanted to know if an *S* in a central position would change his judgment of a relatively nonambiguous stimulus more readily than would an *S* in a peripheral position when faced with different judgments by fellow group members. If so, we would have reason to believe that proneness to change (tendency to conform) is related to degree of vulnerability to direct pressure. If the opposite should occur, i.e., if *Ss* in central positions changed less readily than *Ss* in peripheral positions, we would have reason to believe that central *Ss* were using their favorable position to get others to agree with them. Also, analysis of message content would reveal this process if it should occur.

Method. Since we were interested primarily in comparing central and peripheral positions, we selected the net in which this differentiation is greatest, namely, the star shown in Fig. 1. The task was to judge the number of clicks which were presented to *Ss* via tape recorders and earphones. Each cubicle was supplied with a headset, and a selector panel was set up so that each headset could be connected with either of two tape recorders. With this arrangement, all *Ss* could be given the same number of clicks, or the number given could be different for different *Ss*.

Three conditions were used. In Condition I the central person heard one number of clicks and the other three *Ss* heard a different number of clicks; in Condition

II the central *S* and one peripheral *S* heard one number of clicks and the other two *Ss* heard a different number of clicks; and in Condition III one of the peripheral *Ss* heard one number of clicks and all others heard a different number.

The *Ss* were 96 high school girls ranging in age from 17 to 19 years. *Ss* were divided into 24 groups of four. All groups went through all conditions, but in counter-balanced order.

The clicks had the following characteristics: duration, 5 msec.; rate, 4 per sec.; and intensity approximately 55 db SPL. The number of clicks given on any one trial varied from four to nine. It has been shown (7) that under these conditions clicks can be counted with great accuracy (better than 95% correct). Discrepancies created by sending different numbers of clicks to different *Ss* were either one or two. Following a familiarization trial in which all *Ss* heard the same number of clicks, each group was given two trials in each condition, one with a discrepancy of one and the other with a discrepancy of two. No differences due to discrepancy were observed; therefore, the results from the two trials were combined for purposes of analysis.

Each *S* was required to make an independent estimate of the number of clicks heard on each trial after discussing it with others in the group via written messages. At the end of the experimental session each *S* was required to rate her satisfaction during each condition. On the rating scale, conditions were identified simply as the first one-third, the middle one-third, and the last one-third of the experimental session.

Results. The results of Experiment II are given in Table 3. These results were considered both in terms of the group as a whole, and in terms of different treatments of *Ss* within groups. In the first kind of analysis it was possible to isolate the effects due to (a) conditions, (b) order in which conditions were presented to the group, and (c) the interaction of these two variables. In the second kind of analysis, scores for each condition were

TABLE 3
OUTCOME OF EXPERIMENT II

		Condition I		Condition II			Condition III		
		C*	P _{pp}	C _p	P _o	P _p	C _{pp}	P _{op}	P
Change of estimate	Per cent	58.3	9.0	12.5	12.5	27.1	2.0	1.0	64.5
Time (min.) to decide	Mean	3.5	2.9	4.9	3.7	4.0	2.8	2.4	2.9
	SD	1.9	1.7	2.5	2.4	2.1	2.0	1.5	1.8
Message units per channel	Mean	9.1	10.0	12.9	15.7	15.1	8.5	8.6	9.9
	SD	6.6	6.6	5.4	7.9	7.0	4.5	4.2	5.0
Message units received	Mean	30.0	9.1	45.7	13.8	12.5	27.1	7.3	9.0
	SD	19.6	17.8	15.9	7.9	17.4	14.0	10.3	5.0
Ratings of satisfaction	Mean	6.0	6.0	6.1	6.1	5.9	6.4	5.9	5.7
	SD	0.9	0.8	0.8	0.7	0.8	0.6	0.9	1.1

* The capital letters "C" and "P" stand for central position and peripheral position, respectively; the subscripts "c" and "p" indicate position and number of supporters. For example, C indicates a central person with no support, P_{pp} indicates a peripheral person with the support of two other peripheral persons, etc.

analyzed separately since the different conditions determined the treatments of the Ss.

The group analysis revealed significant differences due to conditions for time to reach a decision ($p < .001$) and for number of messages ($p < .001$); differences in frequency of change and ratings of satisfaction in the various conditions were not significant. Orders of condition presentation resulted in significant differences in frequency of change ($p < .01$), time ($p < .001$), and number of messages ($p < .01$), but differences in ratings of satisfaction were not significant. The condition $\times$ order interaction was significant for time scores ($p < .05$).

These results are apparently due to the fact that in Condition II (where the group was split two/two) changes of estimate were more likely to be in the nature of compromises than were the changes in the other two conditions. This procedure, however, required more time and messages. When Condition II came early in the experimental series, Ss were "conditioned" to the compromising procedure, and also became less sure of their ability to count clicks accurately. Thus, more changes occurred when Condition II came in the first than in the second or third part of the experimental period. The significant interaction term with respect to time to reach a decision is due to the fact that this score decreased with practice.

The individual analysis revealed significant differences due to treatments for frequency of change in all three conditions ($p < .01$ to $p < .05$), for message units received in all three conditions ($p < .001$ in each case), for time to decision in Conditions I and II ($p < .001$, $p < .03$, respectively), and for ratings of satisfaction in Condition III ($p < .05$). No differences in message units per channel were found.

In both Condition I and Condition III the deviator changed her estimate much more frequently than did the nondeviators. The central deviator (C) did not differ from the peripheral deviator (P) in this respect. However, the central deviator apparently did spend more time trying to get the others to agree with her than did the peripheral deviator, and the fact that peripheral nondeviators in Condition I (P_{pp}) changed significantly more ($p < .05$) than did their counterpart in Condition III (P_{ep}) indicates that she may have

had some measure of success. The effect of position within the communication net upon frequency of change of estimate is shown more clearly in Condition II in which both the central S and the peripheral S who agreed with her changed significantly less frequently ($p < .05$) than did the others in the group. It appears that the S in the more central position attempted to use her position to get others to agree with her estimate, but if the opposition became too great, she changed as frequently as did Ss in the peripheral positions.⁵

Examination of the percentages given in Table 3 shows also that frequency of change of estimate is directly related to the number of persons who disagree with the estimate of a given S ($p < .001$). This result, of course, agrees with results reported by Asch (1).

The findings with respect to ratings of satisfaction are understandable if it be assumed that the increased satisfaction resulting from a favorable position in the net is counteracted by being in disagreement with others in the group. In Condition III these effects are both in the same direction and the expected significant differences among Ss were found; in Condition I the two influences are in direct opposition and no differences were found; and in Condition II the issue is confused, but slight, insignificant differences in the expected direction were found.

EXPERIMENT III

Purpose. The results of Experiment II are reasonably straightforward, but there are some questions still unanswered. In the first place, perhaps the most interesting finding was that in Condition II both the central S and the peripheral S who agreed with her changed less frequently than did the other two Ss in the group. But in this condition one supporter automatically meant two opposers, and vice versa. Thus, it is impossible to determine whether this finding is due to the fact that the central S had a supporter, or to the fact that she had only two opponents, or merely to the fact that the opposition was not unanimous.

In the second place, pressures exerted by other group members were permitted to vary

⁵ One central S reported to the peripheral Ss that others had agreed with her own estimate, when actually they had not. In this case, all the peripheral Ss changed their estimate to agree with that of the central S.

with the personality of the Ss; consequently, the amount and type of pressure actually exerted on the Ss in different positions might have been affected by factors other than position within the net and/or the support-opposition ratio.

Finally, a trial varied in length so that the observed differences might have been due to differences in the amount of time spent trying to reach agreement. This factor might also have influenced the ratings of satisfaction. Experiment III was designed to alleviate these difficulties.

Method. The subjects for this experiment were male undergraduates at The Johns Hopkins University. Before the beginning of Experiment III, six groups drawn from this population were run under the conditions provided in Experiment II. Since their responses were indistinguishable from those of the high school girls, it was concluded that any differences between the results of Experiment II and Experiment III probably would not be due to population differences.

The communication net was again the star shown in Fig. 1, and the task was to estimate the number of clicks heard via earphones. In this experiment, however, *E* enlisted the aid of three confederates who ostensibly served as Ss but made estimates according to *E*'s instructions. Thus, only one person (*S*) in each group was naive and it was his responses that were of interest. Each confederate was provided with a list of "agreeing" messages (e.g., "I also heard six clicks"), a list of "disagreeing" messages (e.g., "No, I heard five clicks"), and a list of "neutral" messages (e.g., "Sorry, I could not count them"). On critical trials, each confederate waited until an estimate had been received from *S*, and then he sent a message taken from the appropriate list. Further communications from *S* were answered by transmission of the next message on the list. This process continued until the end of the two-minute discussion period. Thus, the content of the messages sent by the confederates was controlled, but the number of messages sent was not. However, the variability within conditions was not great.

Six conditions were investigated involving three degrees of support-opposition in each of the two positions (central and peripheral) within the net. The three support-opposition conditions were designed to isolate the variables which were operating in Condition II of Experiment II. In one condition, one confederate agreed with *S*'s estimate and the other two were neutral. No change of estimate was expected in this condition, but it was included to obtain ratings of satisfaction and measures of message units transmitted. In the second condition, one confederate agreed and two disagreed with *S*, and in the third condition, two confederates disagreed with *S* and one was neutral.

Ten Ss were run under each of the six conditions. Each group was given ten trials. On each trial Ss were allowed two minutes for discussion after listening to the series of clicks and before making their estimates. On six trials, all confederates agreed with *S*, and on four, the critical trials, the confederates followed *E*'s instruc-

TABLE 4
OUTCOME OF EXPERIMENT III

		Central Position			Peripheral Position		
		ANN*	ADD	NDD	ANN	ADD	NDD
Frequency of change	Per cent	0	12.5	55.0	0	12.5	30.0
Magnitude of change	Mean	0	0.9	4.4	0	0.5	2.0
	SD	0	1.4	2.9	0	0.2	2.8
Message units per channel	Mean	8.2	13.2	11.9	7.3	8.2	8.0
	SD	3.6	3.6	3.4	3.3	2.2	3.9
Message units received	Mean	21.5	23.2	24.5	13.8	13.5	16.2
	SD	4.2	3.4	4.1	1.8	1.6	4.8
Ratings of satisfaction	Mean	8.8	8.7	8.6	7.8	7.3	5.8
	SD	0.4	0.7	1.0	1.5	2.2	2.1

* The letters "A," "D," and "N" stand for agreeing *S*, disagreeing *S*, and neutral *S*, respectively. Thus ANN means that one confederate agreed with the naive *S* and two gave no estimate, ADD means one confederate agreed and two disagreed, etc.

tions as described above. The discrepancy created by the confederates disagreeing with *S* was always two. The maximum frequency of change for any one *S* was therefore four and the maximum amount of change (magnitude) was eight.

At the end of the experimental session each *S* was asked to rate his satisfaction with the group experience.

Results. The results of Experiment III are given in Table 4. Although the central *S* changed his estimate in the "two disagree-one neutral" condition more frequently than did the peripheral *S*, the over-all difference in frequency of change for Ss in the two positions was not statistically reliable. The mean magnitude of change was greater for the central than for the peripheral *S* in both pressure conditions, and this difference was significant ($p < .05$). Likewise, the central *S* sent more messages per channel⁶ ($p < .001$), received more messages ($p < .001$), and rated his satisfaction higher ($p < .001$) than did the peripheral *S*.

Statistically reliable differences were also found between the different support-opposition conditions for frequency of change ($p < .05$), magnitude of change ($p < .001$), message units transmitted per channel ($p < .05$), messages received ($p = .05$), and ratings of satisfaction ($p = .05$). None of the interactions was significant.

Discussion. At first glance the finding that the central *S* changed more than did the per-

⁶ Since time was constant for all trials and all Ss, it was believed that mean number of message units per channel would reflect only differences due to the organizational aspect of positions within the net.

ipheral *S* seems to contradict the results obtained in Condition II of the second experiment; both the central *S* and the peripheral *S* who agreed with him changed *less* than did the other group members. When the differences in experimental procedure are remembered, however, the contradiction appears to be resolved. In Experiment II all *Ss* were naive and each actively tried to get others to agree with her estimate, whereas in Experiment III the one naive *S* actively tried to get the confederates to agree but the confederates responded passively. That is, the confederates refused to change, but they also made no attempt to get the naive *S* to agree with their estimate.⁷ In other words, it appears that the central *S* more than the peripheral *S* tries to get others to agree with his estimate, as indicated by the finding that the central *S* sent more messages *per channel* than did the peripheral *S* under similar conditions of support-opposition. If he fails, however, he changes more than does the peripheral *S*, presumably due to the greater amount of pressure exerted by the nonagrees as indicated by the fact that the central *S* also *receives* more messages than does the peripheral *S*. Failure to find a similar difference between the central *S* in Condition I and the peripheral *S* in Condition III of the second experiment appears to be due to the fact that the central *S* was sometimes successful in getting others to agree with his own estimate. When this happened, the central *S* of course did not change his estimate and differences in frequency of change due to differential amounts of pressure were cancelled out.

It is interesting to note that the results of the first two experiments agree with the results obtained by Goldberg (9) in an experimental situation which also permitted change of opinion by all *Ss*. He reported that group members with less distance (the more central members) were influenced less as the group arrived at a decision.

Differences in change of estimate due to the support-opposition variable are in general agreement with those found in Experiments I and II and with the results reported of other investigations (1). Comparing the results of Experiments II and III it appears that having

⁷ The latter type of situation is similar to that labeled "informational social influence" by Deutsch and Gerard (4).

one supporter strengthens the resistance of an *S* relatively more than the mere reduction of the size of the opposition from three to two persons. It is further suggested that having a supporter strengthens the tendency to resist pressure relatively more than does the simple fact that the other group members are not unanimous in their opposition (cf. 1).

The ratings of satisfaction agree with the results of the first two experiments and further support the hypothesis that satisfaction is jointly determined by position within the net and the amount of support received from other group members. However, in the central position, differences in ratings of satisfaction due to differences in amount of support were not great. Apparently, the level of satisfaction was near maximum⁸ in the central position even with no support from other group members.

SUMMARY

Three experiments dealing with the general problem of decision making in communication nets are reported. In the first experiment, groups of four persons were required to solve two human relations problems in three different communication nets: the star, the slash, and the comcon. The results were as follows: (a) *Ss* in the star required significantly more time to reach a decision than did the slash, which required more time than did the comcon; (b) there were no significant differences among nets in ratings of satisfaction nor in number of messages transmitted, although significantly more messages were directed toward deviators than toward nondeviators; (c) although frequencies were too small for statistical evaluation, there were more deviators in the star than in the slash, and more in the slash than in the comcon. Also, there were more deviators in peripheral positions than in more central ones.

In Experiment II we attempted to discover whether a central *S* would change his judgment of a relatively nonambiguous stimulus more readily than a peripheral *S* when faced with different judgments by other group members. Twenty-four groups of four were placed in the star communication net and required to estimate the number of clicks in a series after discussing it with others in the group. Central and peripheral *Ss* did not

⁸ The highest rating possible was nine.

differ in frequency of change when all others disagreed, but when the group was equally divided both the central *S* and the peripheral *S* who agreed with him changed less frequently than did other peripheral *S*s.

The third experiment attempted to isolate the effects of the support-opposition variable upon judgmental change in different positions in the star. Sixty groups composed of one *S* and three confederates (ostensibly *S*s) were run. The task was the same as in the second experiment. Confederates either agreed or disagreed with *S*'s estimate, or made no estimate. The central *S* changed significantly more than did the peripheral *S* and both changed significantly more with increasing opposition.

From the results of these experiments the following conclusions were drawn: (a) The central *S* more than the peripheral *S* tries to change the opinion of those who disagree, but if he fails, he himself changes more, presumably because of his greater vulnerability to direct pressure; (b) the presence of one supporter strengthens the resistance of a *S* relatively more than the mere reduction of size of opposition and more than the simple fact that the opposition is not unanimous; (c) member satisfaction is a joint function of centrality and amount of support by other group members.

REFERENCES

1. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952. Pp. 450-501.
2. CHRISTIE, L. S., LUCE, R. D., & MACY, J., JR. *Communication and learning in task-oriented groups*. Cambridge, Mass.: Research Laboratory of Electronics, 1952.

3. DEUTSCH, M. An experimental study of the effects of co-operation and competition upon group process. *Hum. Relat.*, 1949, **2**, 199-231.
4. DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629-636.
5. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92-99.
6. FESTINGER, L., & HUTTE, H. A. An experimental investigation of the effect of unstable interpersonal relations in a group. *J. abnorm. soc. Psychol.*, 1954, **49**, 513-522.
7. GARNER, W. R. The accuracy of counting repeated short tones. *J. exp. Psychol.*, 1951, **41**, 310-316.
8. GILCHRIST, J. C., SHAW, M. E., & WALKER, L. C. Some effects of unequal distribution of information in a wheel group structure. *J. abnorm. soc. Psychol.*, 1954, **49**, 554-556.
9. GOLDBERG, S. C. Influence and leadership as a function of group structure. *J. abnorm. soc. Psychol.*, 1955, **51**, 119-122.
10. LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38-50.
11. SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190-207.
12. SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *J. abnorm. soc. Psychol.*, 1954, **49**, 547-553.
13. SHAW, M. E. Group structure and the behavior of individuals in small groups. *J. Psychol.*, 1954, **38**, 139-149.
14. SHAW, M. E. Some effects of problem complexity upon problem solution efficiency in different communication nets. *J. exp. Psychol.*, 1954, **48**, 211-217.
15. SHAW, M. E. A comparison of two types of leadership in various communication nets. *J. abnorm. soc. Psychol.*, 1955, **50**, 127-134.

Received May 31, 1956.

A CROSS-CULTURAL STUDY OF PERCEPTUAL PREDOMINANCE IN BINOCULAR RIVALRY¹

JAMES W. BAGBY

Roosevelt Hospital, New York, New York

ONLY recently has research on binocular rivalry and fusion begun to explore the influence that objects which possess particular significance or meaning for an observer have upon him. Most former work had used abstract figures, circles, squares, colored patches, discrepant lines and the like. Consequently the principles of binocular fusion were elaborated on the basis of studies employing such abstract forms. A recent series of investigations by Engel (3), which employed photographs of people and objects, has demonstrated that there are important empirical and theoretical consequences when "meaningful" content is introduced in fusion and rivalry experimentation. This work strongly suggests that meaning for the perceiver is influential in the binocular resolutions achieved.

Binocular presentation of materials of disparate content and of varying subjective significance thus appears to be a technique of considerable value for psychology (1, 2, 4). The binocular conflict method offers an approach for ascertaining the role of a variety of factors in the perceptual processes. The nature of the specific resolution of two affectively charged pictures which Engel (3) presented to his subjects, for example, appeared to derive from an unconscious and spontaneous choice on the part of the subject. Although a large number of resolutions were possible, the perceptual choices were shown to reflect the subject's actual feelings and dispositions. There is need, however, for much further work to clarify the phenomena encountered in such situations of binocular conflict.

The present investigation is concerned with discovering whether the cultural characteristics of conflicting visual presentations are differentially perceived by members of different societies. Presented with a situation of simultaneous binocular rivalry, do subjects perceive

more readily and consistently visual presentations of content drawn from their own culture than presentations of similar content from another society? The present study investigated this question with respect to Mexicans and Americans.

METHOD

Subjects

There were two experimental groups, one composed of 12 Mexican Ss, and the other composed of 12 American Ss. Half of each experimental group were males and the other half females. They ranged in age from 16 to 42 years. All of the Ss had 20-20 vision. Each S in the Mexican group was matched with an American counterpart as to sex, age, education, occupation and socioeconomic status. Table 1 presents some characteristics of the matched Ss.

An examination of Table 1 reveals that the Mexican Ss were largely middle-class residents of interior provincial centers and that they possessed superior educational backgrounds. The American Ss were residents from Northeast, Middle Atlantic, and Midwestern states with similar educational and socioeconomic backgrounds. With the exception of one matched pair, no S had traveled outside his own country. Thus, for both Mexican and American Ss, the knowledge of the opposite culture was limited to books, the mass communication media, secondhand experience, and other similar sources.

Stimulus Material

Ten pairs of photographic slides were presented. In each pair, one photograph was of a typical Mexican scene including one or more persons, and the other was a photograph of a similar American scene. The picture had been reduced or enlarged to occupy a standard 2" by 2" area. An attempt was made to attain similarity in the form, contour of major mass, texture, definition, and light and shadow of the paired scenes.

The following were the 10 stereogram slide pairs:

- I. An American business man and a Mexican peasant.
- II. A blonde American girl and a dark Mexican girl.
- III. An old farm woman and an Indian peasant woman.
- IV. An old farm man and an old Mexican peon.
- V. An American miner and a Latin American miner.
- VI. An American wedding scene of a bride and groom and a traditional Mexican wedding scene.
- VII. An American farm couple and an Indian couple.
- VIII. A blonde American mother with children and a dark Indian mother with children.
- IX. A baseball scene and a bullfight scene.

¹ This research was made possible through a Rockefeller Foundation Grant to Princeton University for perceptual research. The writer, a research consultant to the Department of Psychology, conducted this project in Mexico during the summer of 1955.

TABLE 1
SOME CHARACTERISTICS OF MEXICAN AND AMERICAN Ss

Male Ss	Age	Education†	Occupation	SES‡	Geographic Residence
A	32	17	Statistician	Middle	Chihuahua
A'	30	16	Accountant	Middle	New York
B	19	13	Student	Middle	Durango
B'	18	13	Student	Middle	Connecticut
C	20	15	English Student	Upper	Coahuila
C'	20	15	English Student	Upper	Wisconsin
D	19	13	Mathematics Student	Middle	Jalisco
D'	18	13	Mathematics Student	Middle	Pennsylvania
E	20	14	Education Student	Middle	Morelia
E'	20	14	Education Student	Middle	New York
F	23	12	Fine Artist	Upper	Mexico City
F'	22	13	Commercial Artist	Upper	Massachusetts
Female Ss					
G	27	17	College Teacher	Upper	Torreon
G'	27	16	High School Teacher	Middle	Kentucky
H	42	16	Language Teacher	Middle	Coahuila
H'	40	17	Language Teacher	Middle	New England
I	33	16	Elementary Teacher	Middle	Chihuahua
I'	32	16	Elementary Teacher	Middle	Illinois
J	16	11	Student	Upper	Guadalajara
J'	17	12	Student	Upper	New Jersey
K	27	12	Artist	Middle	Jalisco
K'	30	16	Art Teacher	Middle	Iowa
L	29	16	Teacher	Middle	Guanajuato
L'	31	16	Teacher	Middle	Colorado

* Ss matched: A, Mexican—A', American; B, Mexican—B', American, etc.

† Numeral refers to years of formal education. Thus, 11 is equivalent to junior year in high school, 17 is equivalent to one year beyond college graduation, etc.

‡ SES refers to socioeconomic status in three groupings: upper, middle and lower classes.



—Reprinted by special permission of the SATURDAY EVENING POST. Copyright © 1955 by the Curtis Publishing Company.



—Reprinted by special permission of Alfred A. Knopf. Copyright by Manuel Gonzales, photographer.

FIG. 1. EXAMPLE OF STIMULUS PAIR: STEREOGRAM IX

X. A young American boy and a Mexican boy.
Pair IX is reproduced in Fig. 1.

Apparatus

The stereogram slide pairs were presented to *S* by means of a prism lens stereoscope enclosed in a light-tight box, like that more fully described by Bagby and Engel (2, 4). Each eye was independently presented with but one photograph of a pair. The slide holder

was adjustable to permit the proper setting for a normal binocular fusion of the two stimuli. The intensity of illumination upon either slide could be independently varied, but for the present set of data it was equated.

Procedure

The *S* was informed that he was to view a series of slides in the apparatus and was to tell *E* what he saw. The *S* was cautioned to look with both eyes and to keep

his forehead pressed against the metal view piece. When the preliminary questions regarding procedure were answered to *S*'s satisfaction he was placed in viewing position.

The initial presentation was stereoscopic slide 33 of the Titchener series (7). The *S* was asked to adjust the setting so the white circle (in one visual field) was centered as nearly as possible in the black circle (presented to the other field). When this had been achieved a slide from the series of Engel's "male faces" (3) was presented and *S* was asked what he saw. All described a single male face. The *S* was then asked if the face appeared in clear focus. Ordinarily there was some minor adjustment to be made in the setting of the apparatus to obtain proper fusion.

At this point *S* was told that the apparatus was adjusted for his own vision and that the experiment *per se* was to begin. The sequence of Mexican-American stereogram slides was then presented, each being exposed for a 60-sec. viewing period. To control for eye dominance, the left-eye-right-eye positions of the Mexican and American scenes were randomized within the series. In terms of *S*'s running descriptive report of what he saw, *E* recorded the relative predominance of the right or left stimulus object. The basic data, however, derive from the initial view or first 15-sec. exposure of the slides.

RESULTS AND DISCUSSION

A variety of reports was obtained from *S*s:

1. In the simplest (and relatively infrequent) case *S* reported seeing but one slide of a pair, and solely described its features. This slide could readily be classified as "predominant" or "preferred" by *S*.

2. In other cases one slide of the pair would be described very thoroughly, after which *S* would begin to say he saw "something else," gradually describing the features of the other slide while reporting that each picture "comes and goes." In the case of such "rivalry," the preference or predominance of one slide was determined by: (a) the picture first reported, and (b) *S*'s statements as to which picture seemed to be present most of the time.

3. There were also reports of a single picture which was an "admixture" of the two pictures presented. "Admixture" reports were rare for the initial view; usually they came during total exposure after a transition from marked predominance. It was not difficult to determine the preferred slide since *S*'s report of content was definitive.

4. There were also reports of the two pictures being seen at once with one superimposed on the other. This condition was reported for first view on only three occasions.

Table 2 gives the over-all results. The Wil-

TABLE 2
DIFFERENTIAL SELECTION OF AMERICAN CONTENT IN
TEN STEREOGRAMS

Nationality	Number of <i>S</i> s Selecting American Content in given Number of Stereograms				Total
	8-10	5-7	2-4	0-1	
Mexican <i>S</i> s	0	0	5	7	12
American <i>S</i> s	9	3	0	0	12
Total	9	3	5	7	24

TABLE 3
OVER-ALL PERCEPTUAL PREDOMINANCE IN THE TEN
STEREOGRAM PAIRS

	No. Where Mexicans Dominate	No. Where Americans Dominate	Total (6 × 10)
Mexican males (6)	44	16	60
Mexican females (6)	45	15	60
American males (6)	7	53	60
American females (6)	12	48	60

coxon (8) test was used for determining the significance of differences between the selections made by the Mexican and American *S*s. Significance was at the .01 level of confidence.

Since the initial time interval was rather short, it may be suggested that culturally familiar subject matter was preferentially reported in cases 3 and 4 without the selection being strictly perceptual. This possibility was checked by excluding these cases from the analysis; the over-all results remain unchanged with significance at the .02 level.

An additional check on the preferential reports for cases 3 and 4 was conducted on all *S*s using the same stimulus materials. This consisted of re-exposing the materials and varying the intensity of light on the two pictures (2, pp. 10-11). The degree to which illumination of the less preferred picture had to be increased to offset the advantage of the other provided a quantitative measure of perceptual predominance. The direction only, and not the relative strength of the indexes, was used in further support of the predominance report derived from the first-view-in-time. In no instance were the original criteria for determining preference, more especially cases 3 and 4, contra-indicated.

Table 3 gives the results from the intensity variation.

Since the probability of equal perceptual predominance was rejected at the .01 level for both *S*s and content, the conclusion seems

justified that Ss report scenes of their own culture as predominant in binocular rivalry over scenes from another culture. The national cultural differences appear critical in affecting perceptual predominance in the majority of the stereogram slide pairs. There were only three pairs in which cultural preference was not marked (i.e., I, V, and X). One can only speculate concerning the less marked cultural influence in the latter pairs, which could involve the physical characteristics of the photographs (e.g., greater clarity or better definition of the prints), or possibly greater "appeal" of the particular individuals depicted regardless of cultural setting. However, the major finding is clear.

Various theoretical interpretations are possible for describing the findings of the present investigation. The position here favored, however, is that advanced by Ames, Cantril, and the transactional school who regard perception as being fundamentally determined by previous, rather than present, experience (5, 6). In transactional perceptual theory the role of meaning is accorded a central position in the perceptual processes. Differences in ways of perceiving come about as a consequence of differences in past experiences and purposes. These in turn emerge from influences in the home, in the school, and in the various groups with which an individual identifies. Thus, under conditions of perceptual conflict as found in the binocular rivalry situation, those impingements possessing the more immediate first-person meaning would be expected to predominate in visual awareness. The findings of the present experiment seem accountable in these terms. While greater familiarity with the objects and scenes from the S's own culture provided a more dominant set for perceiving those materials in preference to the relatively less familiar, other variables such as mood, need, and self-reference values undoubtedly entered into the actual selections. The best

explanation of the present data, therefore, would seem to be in terms of past experience. The accumulated past experience of the individual within the characteristic settings and with typical individuals of his own culture should make pictorial scenes from his own culture possess greater personal significance at the visual level than that possessed by the less familiar situations and individuals of another society.

SUMMARY

Twelve Mexican and twelve matched American subjects were simultaneously presented with a series of ten stereogram slide pairs of similar scenes, one from Mexico and one from the United States. Under these conditions of experimentally induced binocular rivalry, it was found that scenes from the S's own culture tended to be perceptually predominant. The findings were accounted for in terms of the demonstrable role of personal significance in perceptual processes.

REFERENCES

1. BAGBY, J. W. The relative roles of information and action in the genesis of a perception. Doctor's dissertation, Columbia Univer., 1955.
2. BAGBY, J. W. A perceptual study of cross-cultural attitudes in Mexico and the United States. Advance report, Psychol. Dept., Princeton Univer., March 1956.
3. ENGEL, E. Binocular conflict and resolution. Doctor's dissertation, Princeton Univer. 1955.
4. ENGEL, E. The role of content in binocular resolution. *Amer. J. Psychol.*, 1956, **69**, 87-91.
5. ITTELSON, W., & CANTRIL, H. *Perception: a transactional approach*. New York: Doubleday, 1954.
6. KILPATRICK, F. P. Human behavior from the transactional point of view. Hanover, N. H.: Institute for Associated Research, 1952.
7. Titchener Series. *Stereoscopic slides*. Chicago: C. H. Stoelting Company, 1942.
8. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Company, 1949.

Received June 8, 1956.

PSYCHOGENIC DEAFNESS AND PERCEPTUAL DEFENSE¹

MILTON L. KLEINMAN²

New Jersey State Diagnostic Center, Menlo Park

FROM investigations of the relationship between motivation and perception has come the concept of perceptual defense (1, 8, 10), an adaptive process postulated to account for the raised perceptual thresholds obtained when subjects are confronted with unfavorable or anxiety-evoking stimuli. There has been an attempt, however, to explain these increased thresholds on the basis of the relative frequency of the stimulus words used, and on the reluctance of subjects to repeat aloud stimulus words that may be considered socially taboo (3, 4, 14).

The aim of the present study is to investigate whether characteristic defensive reactions to anxiety-evoking stimuli may be demonstrated through changes in auditory perceptual thresholds. It is hypothesized (a) that perceptual defense against anxiety-producing, verbal stimuli can serve to elevate auditory thresholds, and (b) that individuals whose auditory thresholds are characteristically raised in their daily life as a defense against anxiety-provoking stimuli can be identified on the basis of their perceptual behavior in the laboratory.

In order to test these hypotheses the following experimental criteria must be satisfied:

1. The experimental subjects must have similar modes of reacting to inimical auditory stimuli.
2. All stimulus words used must be a part of socially approved everyday speech so that there can not be any inhibition against their use in the experimental situation.
3. The incidence of occurrence in the language of the stimulus words must be carefully controlled.

¹ An abridgment of a dissertation in the Department of Psychology, submitted to the faculty of the Graduate School of Arts and Science, New York University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Degree conferred in June 1953.

² The author wishes to express his sincere appreciation to Professor Leland W. Crafts for his guidance throughout the course of this investigation. He is also indebted to Dr. Moe Bergman who was especially helpful in preparing the recordings. This research was conducted at the Audiology Clinic, New York Regional Office of the Veterans Administration.

4. The auditory intelligibility of the stimulus words must be equated so far as possible.³

METHOD

To satisfy the first criterion the experimental group consisted of 22 Ss with the diagnosis of psychogenic deafness, whose auditory acuity fluctuated without any observable organic basis. It is widely believed that this condition occurs when hearing becomes involved in a basic emotional conflict that the patient desires to repress (7, 11, 15). In terms of personality structure, psychogenic patients have been described as immature, dependent persons who have difficulty with expressing themselves in important interpersonal relationships (12). Although highly dependent upon authority figures, they are unable to recognize this dependency.

The control group consisted of 22 Ss with organic hearing loss who showed no inconsistencies in their ability to hear. All Ss were male veterans of World War II who were patients of the Audiology Clinic at the New York Regional Office of the Veterans Administration. The differential diagnosis was made after a thorough evaluation by the clinic staff which consisted of otologists, clinical audiologists, teachers of speech, teachers of speech (lip) reading, a psychiatrist, a social worker, and a clinical psychologist.

In order to obtain socially approved stimulus words, current literature was searched for a passage of potential interest to World War II veterans involving a sharp conflict with authority. Since conflict with severe authority in relation to deep dependency needs is thought to characterize the psychogenic patients, such a passage might be expected to arouse their anxiety. A passage meeting these specifications was found in the novel by Norman Mailer, *The Naked and the Dead*. The scene in which the homosexually inclined General orders his aide to pick up his cigarette was first adapted for dramatic presentation⁴ and then recorded by professional actors. A count of the important words revealed eleven occurring one or more times. To these were added the words "angry," "force," and "father," which were considered easy associations to the playlet by three clinical psychologists who listened to the final recorded version. These fourteen words formed the crucial or anxiety-evoking list (Table 1). An equal number of unrelated words matched for rhythm of syllabication, accent, and incidence of occurrence in the language according to the Thorndike-Lorge list (14) formed the neutral list.

Since two words reaching the ears at the same level of sound intensity are not necessarily equal in intelligibility, a correction was needed. In a preliminary experiment an intelligibility score, designed to produce homogeneous intelligibility was obtained for each of the

³ The bases for the development of these criteria can be found in the original dissertation, p. 9 f.

⁴ This text appears on p. 36 f. of the dissertation.

TABLE 1
LIST OF CRUCIAL AND NEUTRAL WORDS

Crucial	f*	I.S.†	Neutral	f*	I.S.†
Robert	49	9	Joseph	38	6
power	AA	3	dollar	AA	6
resistance	26	0	delightful	32	3
superior	A	0	immediate	A	6
cigarette	22	3	chocolate	20	0
obedience	15	0	recovery	16	3
unfair	8	0	rainfall	7	3
inferior	19	0	phenome- non	18	3
prison	A	12	delight	A	3
ashtray	{37 17}	6	doormat	{AA 18}	6
lecture	28	0	debate	24	6
angry	A	0	empty	A	6
force	AA	18	chair	AA	0
father	AA	3	daughter	AA	3

* Thorndike-Lorge frequency count per million (A = over 50, AA = over 100).

† Intelligibility score expressed in decibels of attenuation needed to produce homogeneous intelligibility.

stimulus words by a modification of the method described by Harris (2). These are also shown in Table 1.

To prepare the stimulus words for presentation to the Ss, they were first recorded on a disc in random order, by a trained speaker who monitored his voice with a sound level meter. Each word was preceded by the carrier phrase, "The next word is . . ." Each of these words was then re-recorded from the original ten times with a five-second interval between each repetition. During this re-recording, slight variations in peak intensity level due to human errors in monitoring were overcome electronically. The words were then re-recorded again with an attenuator in the circuit to include the correction for intelligibility. To eliminate relative radial distortion and surface wear, the final recordings were made on tape. The stimulus words were arranged in two different random orders, one for the initial series of threshold determinations, and one for the final series. Since they all were re-recorded from the same monitored original disc, each of the stimulus words and its repetitions was identical.

Apparatus

The test chamber was a soundproof room with an ambient noise level of less than 20 db. Facing the S was an Altec-Lansing 604B loud speaker, in a base reflex baffle. There was also a Shure model 556 cardioid microphone to pick up his responses. The S could be observed from the control room through a series of observation windows. This adjoining control room was sound treated, and contained a Jensen 8-in. monitor speaker through which the E could hear S's responses. It also contained the high fidelity sound reproducing equipment, which included a Gates CB7 16-in. transcription turntable with an Audak pickup and transcription arm for playing the recorded playlet, and a Magnecorder recording and playback unit type PT6-AM with a type PT6-J amplifier for playing the tape. The turntable fed to a Langevin preamplifier, which with the tape recorder fed to a Daven attenuator calibrated in 3 db steps.

These, in turn, fed through a Brook 2A3 power amplifier to a Daven output attenuator calibrated in 10 db steps, which fed to the loud speaker in the test chamber. The range of sound intensity extended from -10 db to +98 db, and could be controlled to within a difference of one decibel.

Procedure

Each S sat in the test chamber and faced the E, who sat at the control console in the control room. The S's speech-reception threshold was first obtained by the standard method used in audiology clinics (5). This afforded S an opportunity to become accustomed to the test chamber, provided practice in listening through the equipment, and provided a reference point for the experiment itself. Recorded instructions, to say each word that was heard, were followed by one practice trial. The threshold for each stimulus word was determined by the method of limits (ascending series). Each word was first presented at 9 db below the S's threshold. For every repetition of the word, the intensity was increased 3 db, until S reported it correctly once.

After the recognition thresholds for the twenty-eight stimulus words of the initial series were obtained, S was told he was going to hear an excerpt from a broadcast. A transcription of the playlet was then presented at an intensity level sufficiently above his threshold to be comfortably heard, despite any hearing loss. Upon completion of the recorded playlet, he again heard recorded instructions, and one practice trial preceded the final series of threshold determinations.

RESULTS

The organic and psychogenic Ss were closely alike with respect to age, education, and speech-reception threshold. Mean differences were all insignificant.

The change in threshold for the crucial words and for the neutral words was computed for each S by subtracting the combined threshold on the final series from the combined threshold on the initial series for each set of words. Use of the combined thresholds yielded the total change in threshold for each S, for both word types, and avoided fractions in later computations.

For each of the organic Ss, the combined threshold for both sets of words, crucial and neutral, was higher on the initial series than on the final series. For the psychogenic Ss, the crucial words yielded higher combined thresholds for the initial than for the final series in 17 cases. The remaining 5 Ss had higher combined thresholds on the final series than on the initial series. The neutral words yielded higher combined thresholds for the initial than for the final series for 21 of the 22 Ss.

Group Reductions in Threshold

Organic group. The mean change in threshold for the organic group between the initial and the final series was 89.59 for the crucial words. For the neutral words this was 78.27. The difference between the means was 11.32, which was not significant. The reduction in threshold for the crucial words was greater than the reduction in threshold for the neutral words.

Psychogenic group. The mean change in threshold for the psychogenic group between the initial and the final series was 35.32 for the crucial words. For the neutral words this was 73.23. The difference between the means was -37.91 with $p < .07$. The reduction in threshold for the neutral words was greater than the reduction in threshold for the crucial words.

Group Differences in Mean Threshold Reduction

Comparing the mean differential reduction in threshold for the crucial vs. neutral words for the organic group with the corresponding value for the psychogenic group, a higher-order difference of 49.23 is obtained. The critical ratio of this difference was 6.835, which was significant at the .01 level. A test for the homogeneity of variance was not significant. The difference between the performance of the two groups, therefore, is due entirely to the difference between the means.

DISCUSSION

Since none of the crucial or neutral words used in this experiment could be considered taboo or unfit for polite usage, the reluctance to respond that Howes & Solomon (3) found in McGinnies' (8) experiment was eliminated. Also, because each word was not only identical for the initial and final series, but was also identical for every repetition of itself, any influence of variability among the stimulus words themselves, such as occurred in an earlier experiment of Vanderplas and Blake (16) was also eliminated.

Although there was the opportunity for a practice effect with both sets of words, there was a greater opportunity with the crucial words because of their inclusion in the intervening playlet. The over-all practice effect was significantly greater for the organic than for the psychogenic Ss. In addition, 68 per cent of the organic Ss ($p < .05$) had a greater reduction in threshold for the crucial than for the neu-

TABLE 2
COMPARISON OF THE DIFFERENCES BETWEEN TYPES
OF WORDS AND BETWEEN GROUPS

Type of Hearing Loss	Difference Between Reductions in Threshold for Crucial and Neutral Words			t	(f)
	Mean Reduc- tion	SD	Diff.		
Organic*	11.32	19.9			
Psychogenic*	-37.91	26.1	49.23	6.835†	(1.728)

* $N = 22$.

† $p < .01$.

tral words. On the other hand, the psychogenic Ss had a greater mean reduction in threshold for the neutral than for the crucial words.

The differences between the mean reductions in threshold were statistically significant for six of the psychogenic Ss. In terms of response frequency, all the psychogenic Ss showed a greater reduction in threshold for the neutral than for the crucial words. This negative effect of the increased opportunity for practice with the crucial words may be due to the relationship between the content of the play and deep unresolved conflicts.

The first hypothesis, that perceptual defense against anxiety producing, verbal stimuli can serve to elevate auditory thresholds, is not fully supported. Under the conditions of this experiment a tendency towards a decreased practice effect was found instead. This result was similar to data reported by Postman and Bruner (9) in their study of visual perception under stress. They also found a decreased practice effect for the experimental group, compared with the control group, following the stress situation. The present study was unlike that of Postman and Bruner, however, in that in their experiment the effect of the stress situation was general, while in the present experiment the anxiety was restricted to specific words. In this sense the present experiment resembles more closely the study of McGinnies (8), where seven emotional words were dispersed among eleven neutral words, with results consistent with those reported here.

The significant differential performance of the two groups supports the second hypothesis, that individuals whose auditory thresholds are characteristically raised as a defense against anxiety provoking stimuli can be identified on

the basis of their perceptual behavior. This finding is consistent with perceptual defense theory. The decreased reduction in threshold for the crucial words by the psychogenic Ss may be related to a perceptually manifested defense against the deeper conflicts to which the crucial words were related.

By a somewhat different technique, Lazarus, Eriksen, and Fonda (6) found that the accuracy of auditory recognition was related to the clinical evaluation of dynamic personality processes. The present results are consistent with these findings. In the present experiment, the objective characteristics of the perceptual stimuli were held constant. The major variable was the inferred characteristic difference in dynamic personality processes of the two groups of Ss. The significant difference between the groups may therefore be attributed to the influence of dynamic personality processes upon auditory perceptual thresholds.

SUMMARY

The purpose of this experiment was to investigate whether perceptual defense against anxiety evoking, verbal stimuli is observable through a change in auditory perceptual thresholds. An experimental group of 22 Ss with psychogenic hearing loss, and a matched control group with organic hearing loss were employed. Crucial and neutral words, matched for rhythm of syllabication, accent, frequency of occurrence in the language and corrected for homogeneous intelligibility, were presented in two random orders. The crucial words were centrally involved in the text of a playlet concerned with a sharp conflict with authority, which was presented by transcription after the auditory thresholds for the stimulus words had been obtained. After Ss had heard the playlet the threshold for each of the stimulus words was again obtained.

The difference between the psychogenic group and the organic group with respect to the differential mean reduction in threshold for the crucial and neutral words was significant ($p < .01$). The results are interpreted as being

generally consistent with perceptual defense theory. The greater perceptual difficulty evidenced by the psychogenic Ss with the crucial words was considered a perceptually manifested defense against deep conflicts to which these words had been related by the playlet.

REFERENCES

1. ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.* 1951, **46**, 557-564.
2. HARRIS, J. A. Some suggestions for speech reception testing. *Arch. Otolaryng.*, Chicago, 1949, **50**, 388-404.
3. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' emotionality and perceptual defense. *Psychol. Rev.*, 1950, **57**, 229-234.
4. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401-410.
5. HUDGINS, C. U., HAWKINS, J. C., KARLIN, J. E., & STEVENS, S. S. The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope*, 1947, **57**, 57-89.
6. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471-482.
7. MARTIN, N. A. Psychogenic deafness. *Ann. Otol. etc.*, St. Louis, 1946, **55**, 81-87.
8. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244-251.
9. POSTMAN, L., & BRUNER, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314-323.
10. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. M. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142-154.
11. RAMSDELL, D. A. The psychology of the hard of hearing and the deafened adult. In H. Davis, (Ed.), *Hearing and deafness*. New York: McGraw-Hill, 1947.
12. SLOTE, W. H. The personality of the psychogenic hard of hearing adult. Unpublished doctor's dissertation, New York Univer., 1951.
13. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256-270.
14. THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words*. New York: Teachers College, Columbia Univer., 1944.
15. TRUEX, E. H., JR. Psychogenic deafness. *Conn. Med. J.*, 1946, **10**, 907-915.
16. VANDERPLAS, J. M., & BLAKE, R. R. Selective sensitization in auditory perception. *J. Pers.*, 1949, **18**, 252-266.

Received June 11, 1956.

EXPECTANCY CHANGES IN SKILL AND CHANCE SITUATIONS¹

E. JERRY PHARES

Kansas State College

ANY theory of personality employing a construct of expectancy must be prepared to state the conditions under which expectancies change. Obviously, one such condition is the situation, or more precisely, the individual's categorization of the situation. Recently, considerable attention has been given to an analysis of situational variables as they affect psychological testing and behavior generally (3, 5, 8, 9, 11, 13, 15). This paper explores the relationship between expectancy changes and situational variables, and makes a preliminary effort toward establishing a classification of situations in terms of their effect on expectancies.

The study was stimulated by Rotter's theoretical formulation of the nature of social learning (12), in which human behavior is conceived as being a product of choices. Which choice and thus which behavior occurs is for Rotter a function of both the value of the reinforcement and the expectancy that a given behavior will be successful for a given situation or class of situations. Situations, experimental or natural, are composed of cues that lead to expectancies for various behavior-reinforcement sequences. The individual's categorization of the situation determines his expectancy that a given behavior will be followed by a given reinforcement. How the individual categorizes a situation depends upon what prior situations he sees as having been the same, the amount of his prior experience in these situations, and his overall, generalized expectancy.

Lasko (7) recognized a distinction between learning in situations where the effects following a behavior are a function of the behavior itself, and learning in situations where the subsequent effect or reinforcement is controlled by someone else according to a prearranged sequence. For purposes of this study situations may accordingly be defined as involving skill or chance. A skill situation is characterized

by the fact that the occurrence of a reinforcement is related by the subject to his performance, while a chance situation is one wherein *S* perceives the task as so difficult that being right or wrong is not so much a function of skill as it is luck or *E*'s whim. The occurrence of reinforcement, then, is beyond *S*'s control.

In the skill situation *S* should perceive his performance as providing a basis for generalization to future performance. Since he is the effective agent in the situation he may generalize from past similar experiences. If he fails on a given trial, there should be a decrement in expectancy of future success because his scores are a function of his skill. Furthermore, his experience in the situation gives him successively more adequate conceptions of what to expect on subsequent trials. Thus, the number of trials in a skill situation should be a potent factor in determining subsequent expectancies. In chance situations, on the other hand, none of these statements should hold.

For testing this formulation concerning the situational determinants of expectancy changes the following major hypothesis was framed: 1. Increments in expectancy following successes and decrements in expectancy following failures are greater in a situation generally categorized by *S* as involving his skill than in a situation generally categorized by him as chance.

Stemming from the major hypothesis were three subsidiary predictions. 2. A greater proportion of "unusual" shifts in expectancy is predicted in the chance situation than in the skill situation. "Unusual" shifts refer to increments in expectancy after failure or decrements after success. In a chance situation, *Ss* may be expected often to invoke the "gambler's fallacy," basing their expectancy upon fallacious reasoning such as, "Well, I've been right three times in a row, so it's about time I was wrong." The result is a decrement in expectancy following success; a decrement based not so much upon the previous trial as upon a fallacious view of the odds for success and failure.

While the major hypothesis concerns incre-

¹ This paper is adapted from a dissertation submitted in partial fulfillment of the requirements for the Ph.D. in the Graduate School of the Ohio State University. A special debt of gratitude goes to the director of this dissertation, Dr. Julian B. Rotter.

ments after success and decrements after failure (the "usual" case), it is further predicted: 3. That increments and decrements in expectancy are greater in skill situations than in chance situations regardless of whether they are "usual" or "unusual." It is also hypothesized: 4. That the frequency or number of shifts is greater in the skill situation.

Several studies (4, 6, 10, 16) have sought to establish particular relationships between goal value and expectancy, obtaining divergent results. Seemingly, different experimental designs result in different categorizations of the situation by *Ss*, which lead, in turn, to different results. A situational analysis in terms of the chance vs. skill dichotomy seems promising as an approach to establishing order in this research area.

METHOD

Seventy-seven female students from introductory psychology courses were used as *Ss*. Two experimental tasks were devised so that each could be structured by instructions as either chance or skill.

Task A (Color). Ten 2-in. by 3-in. sample paint patches of varying shades of gray (referred to as the standard colors) were mounted in random order of lightness to darkness on different colored 4-in. by 6-in. sheets of construction paper, which in turn were placed on a large drawing board in two rows of five cards each. Thirteen of these same color patches reduced in size to

TABLE 1
EXPERIMENTAL GROUPS AND ORDER OF PRESENTATION OF TASKS

Groups	Order of Presentation	
	1st Task	2nd Task
1.	Color (chance instructions)	Length (skill instructions)
2.	Color (skill instructions)	Length (chance instructions)
3.	Length (chance instructions)	Color (skill instructions)
4.	Length (skill instructions)	Color (chance instructions)

TABLE 2
FIXED SEQUENCE OF REINFORCEMENT

Trial												
1	2	3	4	5	6	7	8	9	10	11	12	13
x*	x	—	x	—	—	x	x	x	—	—	—	x

* x refers to a positive reinforcement (correct matching), success.
— refers to a negative reinforcement (incorrect matching), failure.

2 in. by 1 in. (no one shade occurring more than twice) were pasted, separately, on 3-in. by 5-in. white filing cards. Each card's patch matched exactly the color of one on the board.

Task B (Length). Materials for this task consisted of ten ½-in. wide strips of black construction paper ranging from 1 in. to 2 ⅛ in. in length (referred to as the standard lengths). The difference in length between any two adjacent size strips was ⅛ in. Each of the ten strips was pasted, separately, on 3-in. by 5-in. white filing cards, at varying angles. All of the cards were then mounted at differing angles on a large drawing board. Thirteen of these same strips (color other than black, and no one length occurring more than twice) were pasted at varying angles on 2-in. by 2 ½-in. white cards. Each card's strip was exactly the same length as one on the board.

Procedure. Four experimental groups were used to which *Ss* were assigned in rotation according to their order of appearance. The procedure is summarized in Table 1. Both order of presentation of tasks and the manner in which each task was structured were balanced. The tasks required *S* to match the matching colors (or lengths) with the standard colors (lengths) on the boards. The skill instructions described the tasks as being difficult, but ones in which score reflected ability. The *S* was told that correct matchings were possible and that some people did well and others poorly depending on skill. The chance task, on the other hand, was described as being so difficult that it did not differentiate between people of high matching and low matching ability. The *S* was told that no one made consistently correct matchings and that success here had been found to be only a matter of luck. This basic structure was repeated in different ways several times and the distinction between tasks reiterated before matchings began.

Each *S* was read one of four sets of instructions depending upon to which experimental group she was assigned, and was given thirteen trials of matching on each task. The sequence of reinforcement was fixed, regardless of task or *S*, and is shown in Table 2.

Although reinforcement in both situations in fact bore a chance relationship to *S*'s performance, the instructions apparently exerted a strong effect. The *Ss* did not recognize that they were being reinforced similarly in the two situations. Thus, because of initial structure, *Ss* categorized the tasks differently and therefore gave differing interpretations to their successes or failures. Many spontaneous comments attested to the fact that none of the *Ss* questioned the report of failure or success for any given trial even though in reality it may have contradicted their actual performance. Both tasks were difficult and novel enough to lend credence to *E*'s statements.

The measure of expectancy used was a betting technique developed by Castaneda (2). This method correlates .99 with verbal methods, although the betting method yields somewhat higher expectancies (14). Before each matching trial *S* was asked to bet from one to ten chips as to whether she would be correct in her matching. During the betting *E* recorded the number of chips bet each time and also a cumulative total of chips won to lend reality to the situation.

RESULTS AND DISCUSSION

The data for the major hypothesis was analyzed in the following manner: A score was computed for each *S*, in each type of situation by taking the absolute amount of all "usual" increments and decrements in expectancy (bets) and dividing by the number of "usual" shifts. The mean magnitude of "usual" shifts in the skill situation was 1.8 while the corresponding mean magnitude in the chance situation was 1.5. When the *t* test is applied,² the difference between means is highly significant ($t = 2.6, p = .004$).

The mean proportion of "unusual" shifts to total number of shifts was .16 in the skill situation and .25 in the chance situation. The results are in the predicted direction but not statistically significant ($t = 1.5, p = .07$). It was also predicted that expectancy shifts would be greater in the skill than in the chance situation. Here, the amount of *all* increments and decrements were summated without regard to sign and divided by twelve for each *S*. Means of .93 and .79 were thus obtained for the skill and chance situations, respectively, the difference between them being significant in the predicted direction ($t = 2.3, p = .01$). Finally, the frequency of *all* shifts in expectancy was greater in the skill situation (mean frequency = 6.12) than in the chance situation (mean frequency = 5.49). The *t* here was 2.1 with *p* at the .02 level.

Only the hypothesis with respect to proportion of "unusual" shifts failed to be confirmed at an acceptable level of significance; even in this case results in the predicted direction of borderline significance were obtained. The findings thus support the view that categorizing a situation as skill leads *S* to use the results of his past performance in formulating expectancies for future performances. In chance situations, on the other hand, past performance does not provide a basis for generalization to future trials since *S* is not the effective agent in obtaining reinforcements. These results, limited to situations defined in terms of chance and skill, suggest the wider fruitfulness of a situational analysis of expectancy changes. Moreover, the need seems apparent for some sort of representative sampling of situations as urged by Brunswik (1),

² One-tailed tests of the distribution of *t* were used throughout this research.

going beyond the narrow range investigated in the present study.

SUMMARY

This study attempted to demonstrate the effect of situational variables on expectancy changes. From Rotter's social learning theory it was hypothesized that expectancies for future reinforcement show greater changes following reinforcement in a skill situation than in a chance situation. This prediction was based on the interpretation that since obtained scores (reinforcements) in a skill situation are a result of performance, *Ss* should utilize past performances as a basis for future expectancies for reinforcement in such a situation. In a chance situation, obtained scores are not controlled by the *S*, thus providing little basis for generalization to future trials.

Two experimental tasks were designed involving (a) the matching of colors and (b) the matching of lengths of lines. Seventy-seven female *Ss* were used, assigned to four groups in a systematic design that varied order of task and of instructions creating skill or chance situations. Each *S* had 13 trials on each task, and all received the same pattern of positive and negative reinforcement.

The major finding was that the chance situation produced smaller expectancy changes than the skill situation. Moreover, the frequency of expectancy shifts was greater in the skill situation.

REFERENCES

1. BRUNSWIK, E. The probability point of view. In M. H. Marx (Ed.), *Psychological theory*. New York: Macmillan, 1951.
2. CASTANEDA, A. A method for measuring expectancy as conceived within Rotter's social learning theory of personality. Unpublished master's thesis, Ohio State Univer., 1951.
3. CHEIN, I. The environment as a determinant of behavior. *J. social. Psychol.*, 1954, **39**, 115-127.
4. CRANDALL, V. J., SOLOMON, D., & KELLOWAY, R. Effects of objective probability and reinforcement value on expectancy statements and decision time. Paper read at Midwest. Psychol. Ass., 1955.
5. GIBBY, R. G., MILLER, N. R., & WALTER, E. L. The examiner's influence on the Rorschach protocol. *J. consult. Psychol.*, 1953, **17**, 425-428.
6. IRWIN, F. W. Stated expectations as functions of probability and desirability of outcome. *J. Pers.*, 1953, **21**, 329-335.
7. LASKO, A. The development of expectancies under conditions of patterning and differential reinforcement. Unpublished doctor's dissertation, Ohio State Univer., 1952.

8. LORD, E. E. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, **65**, No. 10, 1-34.
9. LUCHINS, A. S. Situational and attitudinal influences on Rorschach responses. *Amer. J. Psychiat.*, 1947, **15**, 283-291.
10. MARKS, R. W. The effect of probability, desirability, and "privilege" on the stated expectations of children. *J. Pers.*, 1951, **19**, 332-351.
11. PHARES, E. J., & ROTTER, J. B. An effect of the situation on psychological testing. *J. consult. Psychol.*, 1956, **20**, 291-293.
12. ROTTER, J. B. *Social learning theory and clinical psychology*. New York: Prentice-Hall, 1954.
13. ROTTER, J. B. The role of the psychological situation in determining the direction of human behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1955. Lincoln: Univer. of Nebraska Press, 1955.
14. ROTTER, J. B., FITZGERALD, B. J., & JOYCE, J. N. A comparison of some objective measures of expectancy. *J. abnorm. soc. Psychol.*, 1954, **49**, 111-114.
15. SARASON, S. B. *The clinical interaction*. New York: Harper, 1954.
16. WORELL, L. The effect of goal value upon expectancy. *J. abnorm. soc. Psychol.* 1956, **53**, 48-53.

Received June 15, 1956.

SEX DIFFERENCES IN SOCIAL MEMORY TASKS

SAM L. WITRYOL AND WALTER A. KAESS

University of Connecticut

IN AN earlier study (2) the present investigators found marked sex differences favoring females in performance upon two social recall tasks: the memory for names-and-faces subtest of the George Washington University Social Intelligence Test (5), and a names-and-faces miniature life situation. The purpose of the present investigation was to analyze sex differences on three names-and-faces memory tasks: (a) the GWU subtest, (b) the miniature life situation, and (c) a new memory-for-names-and-faces picture test developed by the writers. In the previous research (2) the first two tasks had yielded results reflecting small variances and insufficient ceilings. The third instrument was developed to provide greater discrimination by increasing the number of items, including pictures of both sexes, and minimizing spatial cues through altering the picture pose from memory task to subsequent recall test. In contrast with the GWU subtest which requires the correct association of one of a series of pictures with each name, this new test requires the subject to associate one of a series of names with each picture-portrait. A recognition as well as a recall measure is obtained within the same test. The experimental design included a check on the possibility that differences in general intellectual ability might account for observed sex differences in social recall.

METHOD

Subjects

The Ss were 103 males and 69 females in nine experimental groups at the University of Connecticut. The average group size was 19. Since the various tasks were administered at different sessions during the semester, the number of Ss whose performances were available for analysis reflect some attrition because of absences. In general the Ss appear to be representative of American college students as indicated in published norms on selected ability and personality tests.

GWU Memory for Names and Faces. On the George Washington University Memory for Names and Faces subtest (5) the S is asked to memorize the names under 12 picture-portraits of young men. Later the 12 pictures are presented again with 13 additional portraits. Two other untimed subtests intervene between the original presentation and the subsequent recall task. In the present study this interval was filled with the GWU

Sense of Humor subtest and the temporal span was rigorously set at 20 minutes. From the matrix of 25 pictures S is required to associate the correct one of four pictures presented in multiple-choice form with each of the 12 names. Each correct multiple-choice association is scored 1 with a total maximum score of 12.

Miniature Life Situation. Five males and five females in each experimental group were participants in the miniature social situation. Each was assigned an alias by the group experimenter who conducted a very brief "sidewalk interview" for about one minute. In every case it was made certain that the participant S spoke his alias clearly at least once. Later Ss in each experimental group were asked to match the correct name from the 10 aliases listed on the blackboard with the person who had adopted that name for the experiment. The participant Ss were asked to stand in each group in a prescribed random order to be identified from one of the names on the blackboard. Although the 10 participant Ss were different for each of the nine experimental groups, the 10 names used as aliases remained standard for all groups. A score of one was credited for each correct matching. The total maximum score was 10.

KW Memory for Names and Faces. An 11" by 14" photograph containing twenty $1\frac{1}{4}$ " by $1\frac{3}{8}$ " picture-portraits of 10 college males and 10 college females randomly arranged by sex were given to each subject face down. Under each photograph was a printed name. Instructions were:

Turn over the photograph. Study these pictures carefully. You will be asked to identify the names and faces later. You will have five minutes. The recall task will be very difficult so study the names and faces intently.

Exactly five minutes later each S was presented with a second 11" by 14" photograph containing thirty $\frac{7}{8}$ " by $\frac{7}{8}$ " picture-portraits: the original 20 plus 5 more student portraits of each sex. This set of 30 picture-portraits was also randomly arranged by sex, with numbers from 1 to 30 under each picture. The answer sheet accompanying this photograph consisted of 30 items. Each item contained five alternative name choices and a "foil" designation. All the name choices appeared on the first photograph. Instructions were:

You studied 20 name-face combinations. The sheet you have contains 30 pictures. Obviously there are ten faces you have not studied. Five are males; five are females. These are the foils. Turn to page two of the exercise. First, circle the picture numbers of the foils. Pick exactly five males and five females. After you are satisfied with your choice for the foils, identify the remaining photographs. If you are unsure of an identification—guess. Leave no blanks.

Two scores for each S were derived from this test. A total score consisted of the sum of the number of

"foil" and name designations accurately identified. A recognition score was the sum of the "foil" choices marked correctly.

Test Administrations

The Modified Alpha Examination Form 9 and the GWU names-and-faces subtest were administered in different sessions early in the semester. About two months later, the miniature life situation task and the newly developed KW names-and-faces test were administered to the same Ss in one session for each of the nine experimental groups.

RESULTS AND DISCUSSION

For purposes of analysis the data from the nine different experimental sections were combined and treated homogeneously by sex groupings. Statistical analysis did not reveal significant differences between observers and participants in the miniature life situation. This finding was consistent with a similar analysis in a previous investigation (2). Three experimental sections yielded significantly higher scores than the other six on the miniature situation task, but this difference did not seem to be seriously contaminating because the sexes were equally represented in the aberrant sections. There were no significant differences between sections on scores obtained from the GWU and KW tests. The previous research had shown that pedagogical and instructional sets are not significant factors in the scores obtained on two social memory tasks.

The following results are based upon 103 males and 69 females in nine experimental sections:

1. Highly significant sex differences favoring females were obtained on all social memory tasks as demonstrated in Table 1.

2. Female superiority on the GWU and miniature situation social recall tasks was consistent with results from a prior investiga-

tion (2), although the magnitude of the differences, expressed in standard deviation units, was slightly smaller in this study. The percentage of females below the male mean was 30 for the GWU subtest, and 29 for the miniature situation task, almost identical with previous results.

3. The newly developed KW names-and-faces memory test is most sensitive to sex differences as demonstrated by the high t values in Table 1. The percentage of females below the male median was 21.7.

4. The mean age of the males was 19.88 with an SD of 2.34; for the females, 18.61 with an SD of 1.03. The t value of 4.26 suggested age as an influencing variable, but when age was correlated with total scores on the KW, the obtained product-moment coefficients were .102 and .139 for males and females, respectively. The null hypothesis is not rejected with these obtained correlation coefficients. The effect of the direction of these correlations, however, is to reduce rather than to increase sex differences.

5. The mean scores on the Modified Alpha Examination Form 9 were 108.29 for males and 108.19 for females. The correlation coefficients obtained between this intelligence test and the KW names-and-faces test were .01 and -.02 for males and females, respectively. Thus, intelligence test scores did not discriminate between the sexes, nor were they related to the newly developed KW instrument.

6. The suggestion from unpublished findings in the earlier study (2) that males remember males better than they remember females, and females remember females better than they remember males, was experimentally confirmed on the KW test in this investigation. Evidence for this tendency favoring like-sex recall is presented in Table 2. Standard scores with a mean of 50 and a standard deviation of 10 were employed in Table 2 for facility of comparison. Scores for the 69 males were combined with a random sample of 69 female scores. This population, thus equally distributed by sex, was used to derive transformation equations for male and female photographs on the KW, and for male and female participant-Ss in the miniature life situation. These transformations were then applied to the entire population. Standard score comparisons became necessary because Ss of both sexes found male pictures an easier task than female pictures in the KW, al-

TABLE 1

SUMMARY OF SEX DIFFERENCES ON THREE SOCIAL MEMORY TASKS

Tasks	Males ($N = 103$)		Females ($N = 69$)		t
	M	SD	M	SD	
KW Total	19.87	5.19	24.06	4.89	5.37***
KW Recognition	6.95	1.68	8.22	1.56	5.08***
GWU	9.31	2.42	10.01	1.92	2.11*
Life Situation	7.18	2.68	8.35	2.09	3.21***

* .05 level of confidence.

*** .001 level of confidence.

TABLE 2
ANALYSIS OF STANDARD SCORES BY SEX IN RECALL OF MALE AND FEMALE NAME-FACE COMBINATIONS

Task	Males (N = 103)				Females (N = 69)				t between Male & Female S's
	M	SD	t	r	M	SD	t	r	
KW Male Names-Faces	48.72	10.38	2.77**	.498	52.03	8.52	4.45***	.616	2.28*
KW Female Names-Faces	46.09	8.92			55.98	8.52			7.33***
Situation Male Names-Faces	49.15	10.14	1.42	.517	51.19	9.72	1.93	.315	1.32
Situation Female Names-Faces	47.72	10.75			53.54	7.30			4.23***

* .05 level of confidence.

** .01 level of confidence.

*** .001 level of confidence.

though male and female items were approximately of equal difficulty in the life situation. All relevant comparisons were made easier by this analysis and yielded the same significance levels as obtained by raw score analysis.

In the life situation there is some tendency to remember one's own sex, but only the female Ss approach the .05 level of confidence. It should be noted, however, that the maximum raw score of the same-sex names-faces combinations in the life situation is only five. Hence, this approach is not a sensitive test of the trend considered here. The KW provided a better test of the tendency for same-sex recall because of the greater range in scores. The maximum raw score for same-sex names-faces combinations on this task is 15. In Table 2, probability levels of .01 and .001 for males and females, respectively, are apparent for the tendency to score better in recalling one's own sex rather than the opposite sex on the KW. It is interesting to note in Table 2 that females are superior to males in remembering male and female names-faces combinations on both KW and life situation tasks. This reaffirms the female superiority on total scores for three social memory tasks demonstrated in Table 1.

The large sex differences found on all three social memory tasks in this investigation were not reported in earlier studies (1, 4, 10) with similar tasks. The objective of constructing an instrument more sensitive to sex differences in social memory was successfully attained. The KW names-and-faces test devised by the

present writers resulted in clearly different score distributions between the sexes. The distribution of male scores was bell shaped; the female score distribution was J shaped. While the difficulty level of the KW is appropriate for the males, it is much too easy for the females. An established effect is strongly apparent on the three social memory tasks, and factors related to this effect may be examined. Theoretical explanation of the results, however, must be *post hoc*.

The possibility that a general intellectual factor may account for the obtained sex differences is not likely. Previous investigations (1, 6, 8, 9, 10) have yielded very small to negligible relationships between the GWU names-and-faces subtest and intelligence or a memory factor in intelligence. This general finding is congruent with the present study in which the null hypothesis was not rejected with respect to intelligence, and with our earlier research (2) in which no relationship to digit span memory was obtained. Performance on social memory tasks thus seems to reflect attitudes, interests, and motivational variables: in short, personality characteristics rather than general intellectual functioning. It is likely that a perfect or near perfect score is within the capacity of most of our Ss. Our tentative explanation for the consistent female superiority on the three social memory tasks in this experiment is found in the emphasis upon greater social facility for the female sex role with the concomitant development of better social skills.

The "same-sex recall" tendency in remembering names and faces was a provocative finding on both the miniature life situation and KW picture tasks. One simple explanation might be found in greater exposure to reference names and faces within each sex. The temptation to offer a "deeper" hypothesis must be resisted until more information about individual personality characteristics is available.

The present findings lend some indirect support to Reisman's (7) notion that there has been a continual shift in American character structure from "inner direction" to "other direction" during the past 50 years. In the present study 28 per cent of the males and 32 per cent of the females obtained perfect scores on the GWU names-and-faces subtest, while Moss reported (3) in 1926 that only one person in five hundred from a college population obtained a perfect score on the same subtest three decades earlier. The data are approximately comparable, and they reflect considerable improvement in social memory over a long period of time with two similar student populations.

SUMMARY

The results obtained from three different social memory tasks—one life situation and two names-and-faces picture tests—demonstrated clear superiority for 69 college females over 103 college males. Of the three tasks, a new names-and-faces picture memory test, designed as a more discriminating refinement of the George Washington University Social Intelligence names-and-faces subtest, yielded the largest sex differences. This new instrument, designated the KW Memory for Names and Faces Test, demonstrated a "same-sex recall"

tendency within the population employed; namely, a tendency for males to remember male pictures better than they remember female pictures, and similarly for females to remember female pictures better.

Neither age nor sex differences in intellectual ability as measured by the Modified Alpha Examination Form 9 accounted for obtained sex differences in social recall. Correlations between these intelligence test scores and KW test scores for males and females were essentially zero.

REFERENCES

1. HUNT, T. The measurement of social intelligence. *J. appl. Psychol.*, 1928, **12**, 317-333.
2. KAESS, W. A., & WITRYOL, S. L. Memory for names and faces: A characteristic of social intelligence? *J. appl. Psychol.*, 1955, **39**, 457-462.
3. MOSS, F. A. Do you know how to get along with people? Why some people get ahead in the world while others do not. *Sci. Amer.*, 1926, **135**, 26-27.
4. MOSS, F. A., & HUNT, THELMA. Are you socially intelligent? An analysis of the scores of 7,000 persons on the George Washington University Social Intelligence Test. *Sci. Amer.*, 1927, **137**, 108-110.
5. MOSS, F. A., HUNT, THELMA, OMWAKE, KATHERINE T., & WOODWARD, L. G. *Social Intelligence Test*. (Rev. Form, 2nd ed.). Washington, D. C.: George Washington Univer., 1949.
6. RANDOLPH, JANE M. A statistical analysis of the Moss Social Intelligence Test. Unpublished master's thesis, Univer. of New Mexico, 1934.
7. RIESMAN, D. *The lonely crowd; a study of the changing American character*. New Haven: Yale Univer. Press, 1950.
8. THORNDIKE, R. L. Factor analysis of social and abstract intelligence. *J. educ. Psychol.*, 1936, **27**, 231-233.
9. THORNDIKE, R. L., & STEIN, S. An evaluation of the attempt to measure social intelligence. *Psychol. Bull.*, 1937, **34**, 275-285.
10. WOODROW, H. The common factors in fifty-two mental tests. *Psychometrika*, 1939, **4**, 99-108.

Received June 18, 1956.

METAPHORS FROM MUSIC CRITICISM¹

ROGER W. BROWN, RAYMOND A. LEITER, AND DONALD C. HILDUM

Harvard University

It is axiomatic with modern semantics that the associations between words and meanings are arbitrary rather than appropriate. To be convinced of the truth of this axiom we are advised to contrast equivalent terms from several languages. For some reason writers in English generally hit on the same trio: *horse*, *cheval*, and *Pferd*. Then follows the rhetorical question: "Is there anything about the animal in question that is more *horse*-like than *cheval*-like or *Pferd*-like?"

The only alternative to arbitrary association is usually thought to be phonetic symbolism, the sound that echoes its sense. A recent publication (2) offers evidence that some phonetic symbolism exists in natural languages. However, it is at best a minor linguistic phenomenon and the case for appropriate associations does not rest here. It is based on the nature of language growth. When someone invents a new machine or forms a concept or buys a dog or manufactures a soap powder his first thought is to name it. But these names are almost never arbitrary creations produced by juggling the sounds of the language into a novel sequence. We think hard and ask our friends to help us find a "good" name, a name that is appropriate in that its present meaning suggests the new meaning it is to have.

Sometimes new words are introduced by borrowing words or morphemes from classical languages. The biological sciences have been especially partial to this practice as *photosynthesis*, *streptoneura*, and *margaritifera* testify. In order to savor the appropriateness of these names a classical education is required and so, for most of us, they are functionally arbitrary. A much more common way to make a new name is by using words or morphemes already in the language; either extending the meaning of some word or recombining morphemes. A dog may be called *Spot* or *Rover*, a detergent may be called *Surf*, one kind of personality is said to be *rigid*. *Underwear*, *overcoat*, *railroad train*, *fireplace*, and *phonograph* are compounds

that can easily be analyzed. Language grows by appropriate extension and this appropriateness sometimes enables us to anticipate the meaning of a new word or the extension of an old word.

An especially interesting class of appropriate names is that vocabulary of sensation which is also used to describe people and voice qualities and manners, such words as *warm*, *cold*, *bright*, and *dull*. Asch (1) wondered whether these words were also used to describe people in languages not historically related to English. He found that they were so used in such remote languages as Hausa, Malayalam, and Thai and is inclined to believe that there may be some common quality in thermal and personal warmth which makes the extended use appropriate and not an accident in the history of English. The rhetorical question now becomes: "Is there not something warm about the warm personality and something cold about the cold personality?"

The present research is concerned with the use of sense vocabulary to describe voice quality rather than persons. Music critics often write that a singer's tone is *dry* or *cold* or *white* or *brittle*. These words have familiar nonauditory meanings but their application to tone is sometimes said to be entirely incomprehensible. The critic may fancy that a voice can be *dry* or *white*, but the reader of his criticism has no idea what he means. In elementary composition courses students are sometimes warned against such flowery and uninformative locutions. We decided to work with Ss who were not familiar with the jargon of music criticism; to ask them to apply extended sensory words to recorded voices in order to learn how well they would agree on the denotative use of such terms. As Asch looked for agreement among unrelated languages, we decided to look for agreement among persons. If there should be such agreement we would take it as evidence that there was something "appropriate" in the extension of nonauditory vocabulary to voices. However, we were prepared to doubt that this "appropriateness" lay in some intersensory dimension common

¹ The authors are indebted to the Laboratory of Social Relations, Harvard University, for supporting this research.

to the narrow use of a term and to the voice quality it described. There might be another way to account for metaphorical extension.

METHOD

Selection of the Words

We began by examining the language actually used by music critics for the description of voice quality. We read the reviews of four critics until we had collected for each critic 40 words that were applied to voices and were not simply synonyms for *good* and *bad*. The critics and the dates of their reviews are:

1. Olin Downes, *The New York Times*, December 19, 1952 through April 30, 1954.
2. Irving Kolodin, *The Saturday Review*, April 10, 1954 through December 25, 1954.
3. Winthrop Sargeant, *The New Yorker*, May 22, 1954 through May 21, 1955.
4. George Bernard Shaw, *London Music*, May 28, 1890 through November 19, 1890.

It is amusing to note, in passing, that although Shaw wrote during the Golden Age of Opera when Melba, Patti, and the de Reszkes were singing in London he appears to have listened with a jaundiced ear. The terms most frequently used by the modern critics were nearly all complimentary, but not so with Shaw. He favored such terms as *screaming*, *yelling*, *brawling*, and *goat-bleating*.

We combined the four lists and ranked the words in order of total critical frequency. We then selected the 10 most common antonym pairs whose first dictionary meanings are sensory but not auditory. They are: brilliant-dull, closed-open, coarse-fine, cold-warm, dark-bright, hollow-full, light-heavy, smooth-rough, soft-hard, thick-thin. In addition, we selected the 20 most common words which do not have antonyms applicable to voice quality. Eleven of these words are primarily auditory in meaning and nine are not. The words are: brittle, chromium, dry, gravelly, hoarse, limpid, lustrous, lyric, nasal, pinched, reedy, resonant, ringing, round, shrill, solid, sonorous, strident, velvet, and white.

The Recorded Voices

We selected recordings of nine famous operatic voices; three Italian dramatic sopranos, three tenors, and three baritones. Within a vocal category the voices sang the same fraction of the same operatic aria. Each excerpt lasted approximately 30 seconds. The selections and singers follow:

Sopranos: *Vissi d'arte* from *Tosca* of Puccini. Played through the word *aiutai*.

1. Maria Meneghini Callas (S1).
2. Maria Caniglia (S2).
3. Renata Tebaldi (S3).

Tenors: *Che gelida manina* from *La Bohème* of Puccini. Played through the word *vicina*.

1. Mario del Monaco (T1).
2. Jan Peerce (T2).
3. Richard Tucker (T3).

Baritones: *Eri tu?* from *Un Ballo in Maschera* of Verdi. Played through the words *per me*.

1. Gino Bechi (B1).
2. Lawrence Tibbett (B2).
3. Leonard Warren (B3).

Subjects

The Ss were 47 Harvard undergraduates who proved to be uninformed about opera and music criticism when asked a set of questions concerning these matters. They participated in groups of three, four, or five, but responded independently since all of their answers were written in test booklets.

Procedure

The Ss began with a practice procedure involving a voice and words not used in the experiment proper. In the experiment, the order of presentation of the three voice categories (sopranos, tenors, and baritones) was randomly varied. Within a category, however, the voices were always presented in the same order for reasons of experimental convenience. To begin with, all three voices in a given category were played once. Recordings were played on an Admiral table model phonograph with an attached 12" loudspeaker. Then the first voice was played again and Ss marked the ten antonym pairs for that voice. For each pair there was a scale from one to four extending between the two adjectives, and Ss circled the number best expressing their judgment of the voice in question. It might, for instance, be thought very brilliant—1, moderately brilliant—2, moderately dull—3, or very dull—4. Then the second voice was played and the pairs marked and finally the third voice. After the antonym pairs had been completed for the three voices, Ss listened again to each one and picked the five words from the 20-word check list that seemed most aptly descriptive of the voice in question. The experiment continued in this fashion with the remaining two categories of voice. In summary, Ss were required to judge each voice in terms of each antonym pair and to make free selection of the five most appropriate words from the 20 words listed without antonyms.

Finally, Ss listened to the voices once again and indicated for each one whether or not they liked it. In five cases Ss were privately interviewed after the standard procedure to learn how they would explain their word choices.

RESULTS

The Antonym Pairs

For our first test we have simply dichotomized the choices, combining ratings of 1 or 2 as choices of the adjective on the left, and ratings of 3 or 4 as choices of the adjective on the right. If Ss had no basis at all for deciding which member of a pair to apply to a given voice, they might be expected to choose randomly. In that case the likeliest result for any pair with any voice would be an even division between the two words. A departure from chance expectation significant at the .05 level occurs with a division into 31 and 16 while for the .01 level the division must be 33 and 14, or more extreme. There are 38 differences significant at the .01 level and another 8 at the .05 level. In about half of the cases then, these

words were applied to voices with better than chance uniformity.²

It is possible that subjects favor one word in a pair over the other simply because the word is more familiar or because it is thought to be more suitable to the description of voices of whatever quality. Departure from chance does not demonstrate that the words actually suggest one kind of voice quality rather than another. However, critics use this vocabulary with the intention of suggesting particular qualities. We need to know whether the pattern of choices on a pair for one voice is significantly different from the pattern across the remaining eight voices. If choices on a given pair of adjectives divide in the same way on all nine voices then the words in question do not distinguish one voice from another. The test is chi square. The obtained frequencies of choices 1, 2, 3, and 4 for a particular pair on a given voice were compared with the frequencies to be expected on the basis of the distribution of choices for that pair across the remaining eight singers. Yates' correction was used in every case. A significant chi square indicates that the pair in question is applied to the singer in question in a manner different from the application of the same pair to the other singers. This is a test of the specificity of the semantic rule linking each voice with each adjective pair. It indicates the degree to which these adjectives distinguish one voice from the others. The significant chi squares appear in Table 1. Of the 90 chi squares computed (10 for each voice), 33 are significant at the .05 level or better, which is considerably more than the three or four to be expected by chance.

We also compared each of the three vocal ranges with the other two. Baritones, for example, were contrasted with tenors and sopranos to see if some adjectives showed a different choice pattern for baritones as a group than for tenors and sopranos combined into another group. The test again is chi square. The results make very good sense. Baritones are more likely to be called *dull*,

coarse, *closed*, *dark*, *heavy*, *rough*, *hard*, and *thick*. This is at once a description of baritone quality and of the operatic characters played by baritones. Tenors are especially likely to be called *bright*, *light*, and *thin*, while sopranos are *coarse*, *soft*, *light*, and *thin*. There are more adjectives distinctive for baritones than for either tenors or sopranos and that, presumably, is because the baritone is a low register voice set against two high register voices. Tenors and sopranos are more alike on these dimensions than either is like the baritone. It was conceivable, of course, that the two male voices would resemble one another more closely than either would resemble the sopranos. It is fitting that this should not be so since tenor and soprano on the operatic stage invariably make common cause against the baritone.

The Check List

From 20 words *S* chose 5 to characterize each voice. We wanted to know if a particular word was used more or less often for one voice than for the other eight. The test is again chi square with Yates' correction still necessary. We determined the number of times a particular word was applied to a particular singer and also the number of times it was applied to all other singers. Since there were 47 *Ss*, a word always had the possibility of being applied 47

TABLE 1
THE SIGNIFICANT CHI SQUARES FOR EACH VOICE COMPARED WITH THE EIGHT OTHER VOICES ON THE ANTONYM PAIRS

Antonyms	Voices	χ^2 (<i>N</i> = 47)	Antonyms	Voices	χ^2 (<i>N</i> = 47)
Light-heavy	S1	12.16*	Coarse-fine	B1	26.63*
Thick-thin		29.51†	Cold-warm		40.25*
Coarse-fine	S2	12.67†	Closed-open		8.66*
Brilliant-dull	S3	44.63*	Dark-bright		8.11*
Coarse-fine		35.94†	Hollow-full		10.17*
Cold-warm		34.20†	Smooth-rough		34.62†
Closed-open		24.39†	Soft-hard		17.35†
Dark-bright		33.39†	Coarse-fine	B2	9.96†
Hollow-full		13.78†	Hollow-full		9.77†
Light-heavy		22.86*	Light-heavy		10.92†
Smooth-rough		31.37*	Thick-thin		11.07*
Soft-hard		38.74*	Closed-open	B3	7.89*
Hollow-full	T1	14.35*	Dark-bright		15.32*
Light-heavy		20.25*	Hollow-full		11.01†
Thick-thin		11.37†	Light-heavy		35.28†
Hollow-full	T2	14.33†	Thick-thin		76.05*
Thick-thin	T3	9.25†			

* The first word of the pair was more likely to be applied to the voice in question than to the other voices.

† The second word of the pair was more likely to be applied to the voice in question than to the other voices.

² The complete table of choices for each antonym pair and for each voice has been deposited with the American Documentation Institute. Order Document No. 5136 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.25 for 35 mm. microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 2

THE SIGNIFICANT CHI SQUARES FOR EACH VOICE COMPARED WITH THE EIGHT OTHER VOICES ON THE WORDS FROM THE CHECK LIST

Word	Voice	χ^2	Word	Voice	χ^2
Shrill	S1	42.95*	Shrill	T2	13.59†
Brittle		9.80*	Brittle		5.58†
Round		8.15†	Pinched		6.84†
Solid		15.33†	Round		7.00*
Resonant	S2	8.24†	Solid	B1	8.33*
Shrill		16.38*	Dry		9.61†
Pinched		3.89*	Resonant		4.87*
Sonorous		12.16†	Sonorous		4.06*
Gravelly	S3	5.42*	Brittle	B2	30.04*
Strident		4.90*	Pinched		3.89*
Shrill		4.41†	Round		4.32†
Brittle		6.00†	Dry		18.16*
Pinched	T1	7.11†	Lustrous	B3	8.57†
Dry		12.00†	Chromium		4.43†
Lustrous		10.81*	Lyric		16.42†
Reedy		4.17†	Nasal		11.04*
Sonorous	T2	7.92*	Velvet	B3	10.03†
Chromium		6.92*	Shrill		12.57†
Gravelly		4.93†	Brittle		6.00†
Lyric		22.44*	Round		4.76*
Nasal	T3	10.15†	Solid	B3	18.32*
Strident		7.57†	Shrill		8.01†
Velvet		26.90*	Solid		9.16*
Pinched		8.00*	Reedy		5.94†
Lustrous	T4	6.32†	Resonant		4.29*
Reedy		4.26*			

* The word in question was more likely to be applied to the voice in question than to the other voices.

† The word in question was less likely to be applied to the voice in question than to the other voices.

times to a particular singer and the possibility of being applied 376 (8×47) times to all other singers. The problem was to discover whether the attributions to one singer represented a larger or smaller proportion of the relevant possibilities than did the attributions to all singers. The significant results appear in Table 2. Of 180 chi squares 51 are significant with $p = .05$ or better. This is far more than the six or seven to be expected by chance.

Finally, we contrasted the frequency with which auditory words were chosen and the frequency with which nonauditory words were chosen. The mean number of choices of auditory words is 119 while it is 94 for nonauditory words. This difference does not attain to significance at the .05 level and so it is not clear that Ss found the auditory vocabulary more appropriate to the description of voices than the nonauditory vocabulary. Furthermore, among the auditory words there were no more cases of discriminating use than there were among the nonauditory. The two vocabularies do not differ in the frequencies with which they were used nor in the precision of that use.

DISCUSSION AND FURTHER ANALYSIS

We have asked Ss to apply sensory, non-auditory words to voice qualities. There was high agreement among the Ss on the pairing of some words with some voices. Furthermore, many of these words were used discriminatively; i.e., the pattern of applications to one voice differed from the pattern for other voices. Suppose we knew that our Ss had experience of such words as *white*, and *dry*, and *thick* in their primary application to sense data but not in any extended auditory application. From the fact that the extended use shows agreement and discrimination we might conclude that the primary senses of *white*, and *dry*, and *thick* are concepts having appropriate application to voice quality. In effect, our Ss would have demonstrated that they possess shared, distinctive, supersensory concepts by their agreement on the identification of new auditory instances. The first difficulty with this conclusion is that we have not shown that our Ss have no prior knowledge of the extended auditory use of the words in question.

While it is likely that some of the terms were familiar designators of voice quality it seems likely that most of them were not. To begin with, of course, our Ss were not habitual readers of music criticism and were not knowledgeable about opera and opera singers. That fact alone does not guarantee that they were unfamiliar with the use of this vocabulary to describe voices. We must add that many Ss made complaints about the vocabulary offered them. They said that they would never freely choose such words as *white*, or *chromium* or *brittle* to describe voices. These complaints indicate that much of the vocabulary seemed unconventional to them and that our Ss had to rack their brains to find some reason for calling one voice *closed* and another *open*. The interview protocols confirm this impression. Especially instructive are the observations concerning the words *white* and *open*. In music criticism these terms name undesirable qualities. The *white* tone is colorless; the *open* voice is diffuse, lacking in focus. Many Ss used these terms as names of desirable qualities and interviews disclose how that happened. *White* can stand for purity, for a voice without blemish. *Open* can mean freely and easily produced. In these cases the words functioned as metaphors, as extensions of familiar terms to a new range of experience, extensions that were felt to be appropriate. The metaphor is most apparent

here because it runs counter to conventional critical use. The interviews indicate, however, that with most words *S* was unaware of any conventional usage and felt that he must discover reasons for selecting one word rather than another.

We believe it is reasonably clear that our *Ss* felt themselves to be applying to voices a number of words that would not ordinarily be so applied. Furthermore, *Ss* seem generally to have found one word more appropriate than another and to have agreed on which word it was. In short they appear to have created a number of metaphors and to have agreed on them. How do they come by the ability to perform in this way? Let us mention first the possibility with which this article began, a possibility suggested by Asch's study. It may be that when a child learns the sense meanings of words like *bright*, *dry*, and *rough*, he acquires a set of concepts applicable to voices and persons and manners. There is, however, another possibility that is suggested by factor analysis of our results.

The list of antonym pairs we used includes eight of the pairs used by Osgood and Suci (4) in their factor analysis of linguistic meaning. These experimenters had *Ss* place 20 concepts (lady, boulder, sin, etc.) on 50 descriptive scales (good-bad, large-small, etc.). Analysis of the resultant matrix yielded three orthogonal factors which have been characterized as: (a) evaluation, (b) potency, (c) activity. The same factors have appeared in other studies and, from scales representing these factors, the Semantic Differential was constructed. Our *Ss* placed voices (rather than concepts) on 10 scales defined by adjective antonyms. This procedure is clearly analogous to that of Osgood and Suci and yields data that can be factor analyzed. In addition, our *Ss* applied 20 words (without opposites) to the same voices. These results were also included in the factor analysis. Finally, our *Ss* were asked to say whether they liked or disliked each voice and those judgments have also been included.

The nine voices were ranked along each of the 31 scales according to the number of attributions they received from *Ss*. It seems quite reasonable to assume that the numbers of attributions would be continuously and normally distributed over the universe of singing voices. However, because of the selected range of voices, we could not be sure of normality and it therefore seemed advisable to use only rank

order correlations. In view of the very small *n* (9 voices) and the unorthodox use of rank-order *r* for factor analysis, the following results must be considered very tentative.

The first factor accounts for 56.7 per cent of the common factor variance. It may be called evaluation. The scales having factor loadings above .90 are *cold-warm*, *hollow-full*, and *like-dislike*, and the terms are *brittle*, *dry*, *lustrous*, and *pinched*. It is not possible to make detailed comparisons with the Osgood-Suci evaluation factor because our scales were not all the same. The scales most heavily loaded with their factor included *beautiful-ugly*, *sweet-sour*, *valuable-worthless*, and other pairs excluded from our study because of their obvious synonymy with *good-bad*. There is some indication, however, that the two evaluation factors would not be quite the same even if identical scales had been used. For instance, the scale *wet-dry* in the Osgood-Suci study is not loaded with the evaluation factor whereas the word *dry* is heavily loaded with evaluation in our study. This simply means that dryness is clearly undesirable in a voice but that, across 20 concepts, dryness is neither clearly bad nor clearly good. Changing the material to be judged somewhat alters the character of the evaluation factor but there is justification for calling both factors of evaluation, since a *good-bad* scale is central to each.

The second factor seems to be best characterized as potency. The third is more difficult to characterize, but there is some ground for thinking it an activity factor similar to that found by Osgood and Suci. We extracted a fourth factor, but find it impossible to characterize. It is interesting to note that these four factors account for over 80 per cent of the total variance while the Osgood-Suci factors accounted for only 50 per cent. With more restricted materials and scales we obtained more communality.

What is the origin of the semantic clusters revealed by the Osgood-Suci work and by the present study? Consider two scales from the potency factor; *large-small* and *heavy-light*. Imagine, as is quite possible, that these words were first encountered as names for attributes of a set of wooden blocks. If all the blocks were made of the same kind of wood the two scales would be in natural correlation; the larger the block the heavier it would be. The same sort of relationship binds *heavy-light* to *thick-thin* and to *wide-narrow*. Words of this kind are

ordinarily used to compare one object with others of the same type. A block is *large* relative to other blocks. As a consequence *large*, *heavy*, *thick*, and *wide* are terms that ordinarily imply one another. And the implications spread even more widely. When it is a question of people the larger, heavier, thicker, wider specimens seem to be more likely to have loud bass voices than are the smaller, lighter, thinner, narrower specimens. In short, sense qualities, though logically independent, are probably correlated in nature. Suppose, now, that *S* is asked to judge 20 concepts with respect to all these scales. A concept like "boulder" is referred to rocks and stones and, in comparison, judged to be *heavy*, *large*, *thick*, and *wide*. These terms are directly applicable to boulders. However, boulders have no voices. Where, then, does the concept belong on the *bass-treble* or *loud-soft* scales? We cannot doubt the answer. If Disney were to give a boulder a voice it would be *bass* and *loud* in contrast to the piping of a pebble. This could be a simple extrapolation of the usual correlation between these terms. It does not necessarily mean that there is anything in a bass voice *per se* that makes it appropriate to a boulder.

In our study *Ss* may have felt that *thick* and *thin* simply do not apply to voices. However, *loud* and *resonant* do. Now thick people and animals and violin strings are likely to be *loud* and *resonant*, so if *S* is required to guess he will guess that *loud* and *resonant* voices ought to be called *thick* rather than *thin*. Similarly, *S* may feel that *cold* and *warm*, in their thermal sense, do not apply to voices. However, *warm* things are generally more agreeable than *cold* things and so a voice he likes is going to be called *warm*. Factor analysis suggests, then, that metaphors preserve correlations laid down in the earliest language experience. If *S* is asked to apply a word to a concept or experience to which it has no literal application he will make his decision so as to maintain the usual correlation between this word and others that do have literal application in the present instance. Karwoski, Odbert, and Osgood (3) have suggested that such natural correlations may also help to account for the clustering of sense qualities revealed by the study of synesthesia.

If there is a natural correlation in the use of words like *thick*, *heavy*, and *large* it does not

follow, of course, that these words name the same dimensions. The Osgood-Suci findings must not be interpreted to mean that our lexicon is made up of countless perfect synonyms for *good-bad* and *strong-weak*. We have only to imagine a set of blocks that includes some made of balsa wood as well as some made of pine to realize that *heavy* and *large* name dimensions that are logically independent. However, Osgood and Suci lined up words with concepts to which they had no literal application. Forced to make some decision *S* applied the inapplicable words wherever he applied their usual correlates among the applicable words. So the words are not used with all the independence of which they are capable but, rather, with all the intercorrelation they generally manifest.

SUMMARY

A vocabulary of metaphorical terms used to describe voice qualities was drawn from the writings of four music critics. The *Ss* who were not well informed about music listened to recordings of nine operatic voices and assigned terms from the metaphorical vocabulary to these voices. Many of the words were used with agreement and discrimination. Factor analysis of the results yielded three factors similar to those found in the more general factor analyses of linguistic meaning made by Osgood and Suci. The factors are: (a) evaluation, (b) potency, (c) activity. It is suggested that these factors may derive from correlations in nature between sense qualities. Metaphorical extension of the name of a sense quality may operate so as to preserve such correlations. Metaphorical extension of a sensory term need not be taken as evidence of an intersensory dimension.

REFERENCES

1. ASCH, S. On the use of metaphor in the description of persons. In H. Werner (Ed.), *On expressive language*. Worcester, Mass.: Clark Univer. Press, 1955.
2. BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages. *J. ab-norm. soc. Psychol.*, 1955, 50, 388-393.
3. KARWOSKI, T. F., ODBERT, H. S., & OSGOOD, C. E. Studies in synesthetic thinking: II. The role of form in visual responses to music. *J. gen. Psychol.*, 1942, 26, 190-222.
4. OSGOOD, C. E. & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, 50, 325-338.

Received June 22, 1956.

LIKING FOR THE GROUP AND THE PERCEIVED MERIT OF THE GROUP'S BEHAVIOR¹

LEONARD BERKOWITZ

University of Wisconsin

STUDIES consistently have shown that individuals tend to be more successfully influenced by groups to which they belong the more attractive group membership is to them (e.g., 1, 6). In many of these investigations attraction to the group has been defined in terms of the individual's liking for the other group members, but the findings are generalized to other sources of attraction as well (1). The major purpose of the present paper is to point to an effect of liking for the other group members that has been largely overlooked in theoretical interpretations of the relationship between liking and influence and to suggest that this effect may not occur for other sources of attraction to the group.

Perhaps the most fully developed theory of social influence is that proposed by Festinger (3, 5). Briefly, with regard to group attractiveness, this theory states that individuals are more highly motivated to achieve within-group consensus the more attractive the group is to them because pressures toward uniformity increase directly with this attractiveness (cf. Corollary VII A, 5, p. 131). These pressures supposedly are felt by the individuals themselves as well as being expressed in the behavior of the other members toward them (6).

Liking for the other group members—one source of attraction to the group—also can facilitate the group's influence over the individual members through an additional process. Liking can increase the perceived merit of the group's behavior in the absence of independent objective tests of this merit. Thus, subjects (Ss) in these experiments may have conformed to opinions expressed by the other group members partly because, in liking these others, the Ss perceived their opinions to be probably correct.

Something similar to the present hypothesis may be implied in several statements made by Festinger concerning the individual's acceptance of the group's opinions. He has proposed

¹ The invaluable assistance of Mrs. Marsel Heisel, who collected the present data, is gratefully acknowledged.

(3, 4) that there is a greater tendency for individuals to adopt the opinions advocated by the group the more attractive the group is to them. However, no explanation is given except to suggest that the attractive groups have greater "power" over their members (3), and there is no explicit statement that liking may affect the perceived merit of the other group members' opinions.

The major focus of this study, therefore, concerns the following hypothesis: Ss with high liking for their other group members (in contrast to Ss with lower liking for the others) (a) judge the opinions expressed by the others in the group as being more likely to be correct, and (b) show greater conformity to these opinions.

As a secondary purpose, the present experiment also investigates the effects of past proficiency upon influence. In line with findings of previous studies (9, 10) that Ss are more readily influenced by others who previously have expressed many correct opinions than by those who have often been incorrect, it is reasonable to hypothesize that Ss' evaluations of the merit of others' opinions are directly affected by these others' records of previous correctness on similar issues. One may then inquire whether liking and past proficiency combine to produce effects on the perceived merit of an opinion greater than that resulting from either factor alone, and to yield summative influence effects.

METHOD

Manipulation of Liking

Differences in Ss' liking for their partners were obtained by employing a now standard procedure. At the time of volunteering to participate, students in introductory psychology discussion sections completed brief personality questionnaires on the basis of which they supposedly would be matched with a "congenial" partner. Two Ss, both male or female, were scheduled for a given hour. When each S arrived to keep his appointment, he was ushered into a separate room for oral instructions. The Ss in the *high liking* condition (High Like) were told that the matching had been very successful, that he and his partner should be very congenial and probably would like each other. Each of the

Ss in the *low liking* condition (Low Like) was told that due to scheduling difficulties it was not possible to match him with a congenial partner, and that because the S's personality and that of his partner were so dissimilar they might even grow to dislike each other.

Group Task

Following these instructions, the two Ss were brought together. E then explained that the purpose of the experiment was "to determine how people work together when they have to communicate by indirect means such as over a telephone system." The Ss were led to believe that they would represent an Army artillery team. One of the Ss would be the "observer" and the other the "battery commander." They were shown a series of charts supposedly representing the terrain lying before the "observer" at his forward post. The "observer's" task was to estimate the distance in inches from one point on the chart (his post) to each of a number of other points (the targets). It was explained that the "battery commander" would have two sources of information concerning the distances to the targets, the distance estimates telephoned to him by the "observer," and "an old and possibly unreliable map of the area" from which rough distance estimates were computed to each of the targets. The "battery commander's" task was to compare, for each of the large number of targets, the "observer's" estimate with the estimate from the "old map" and then write down the distance information to be used by his artillery.

Manipulation of Partner Proficiency

After the Ss returned to their individual rooms E told each S that he had been picked to be the "battery commander." The S was given a set of earphones by means of which, it was said, he could hear the "observer." He was also given a list of the 80 targets, the distances to each estimated from the "old map," and a sheet of paper on which he was to write his decisions. Each S actually worked alone and heard a previously recorded tape of the supposed interchange between E and the "observer." No S reported being aware that he was listening to a tape recording. The first part of the experiment consisted of a preliminary series of ten supposed practice trials for the "observer" which was employed in manipulating Ss' perceptions of the task proficiency of their "partners." The Ss were instructed that in the practice trials as throughout the entire task an estimate would be counted as a "miss" if it departed from the true distance by more than one inch. For each of the ten practice targets the supposed "observer's" voice reported a distance, after which the E's voice was heard to say either "hit" or "miss." For the *high observer proficiency* condition (High Prof.), seven out of the ten "observer" responses were said to be "hits," while only three of the ten were termed "hits" in the *low observer proficiency* condition (Low Prof.).

While other evidence will be presented that the manipulation of the "partner's" perceived proficiency was at least partly successful, analyses of variance indicate that it was not completely so. The main effect for proficiency fell short of desirable standards of significance for Ss' judgments of the percentage of hits

(correct opinions) by their "partners" in the main experimental series ($F = 3.18, p < .10$), and was non-significant for the behavioral conformity measure.

Procedure

When the preliminary series was completed, the main experimental procedure was followed for fifty distance estimates in five distinct series of ten targets each (Block 1). For each of these targets, S was to decide upon a distance to be written down using the two types of information supplied to him. The two distance estimates for each target ranged from being in complete agreement to four inches apart.

After the fiftieth target the E's voice asked S to write down the percentage of targets on which he believed the "observer" had scored a hit. The procedure then was repeated for three more series of ten targets each (Block 2), and Ss again estimated the "partner's" percentage of hits. There were no reliable differences among the conditions in degree of change in the estimates from Block 1 to Block 2.

At the completion of the last series the tape recorder was stopped and the Ss were given a brief questionnaire. Three items, proven to be sensitive to the manipulation of liking in other experiments, were used as a scale to assess the liking of the Ss for their "partners." The items are as follows: 1. How well did you like your partner? 2. If you were taking part in another experiment, how much would you like to work with this same partner? 3. If it involved any effort on your part, would you care to know your partner better?

Conformity Measures

The distance decisions made by the Ss as "battery commanders" fell somewhere between the supposed "partner's" estimates and those derived from the "old map." Two conformity scores were arrived at by counting within each of the two blocks of series the total discrepancy (in inches) between the S's decisions and the "partner's" estimates and then dividing each of these sums by the total possible discrepancy in the particular block, that between the values supplied by the "map" and the "observer." Both sets of percentages then were normalized by means of the arc sine transformation. The smaller the scores the more the S's decisions had been influenced by his "partner."

Subjects

Analyses of the results with the scale measuring the Ss' liking for their partners indicated two sources of potential difficulty. First, the difference between the high and low liking conditions on this scale just fell short of significance (p approximately .10). Second, and perhaps more important, the Ss liked their partners more in the High Prof. condition than in the Low Prof. condition. As has been demonstrated in other studies (2, 8), individuals tend to be more attractive to other group members if they are seen as performing well than if they are seen as performing poorly. Since it was important for the purposes of the present experiment to eliminate this last difference, Ss were selected for inclusion in the final sample in such a manner that the liking scores were comparable in the High and Low Prof. conditions but divergent in the High and Low

Like conditions.² This selection resulted in statistically significant differences between the High and Low Like conditions ($p = .01$).

As a result of this screening, and where only the liking scores were known, twelve of the original 52 Ss were eliminated leaving ten Ss in each of the four conditions. Thirteen of the remaining Ss were females who were distributed approximately evenly among the conditions.

RESULTS

Perceived Correctness of the "Partner's" Opinions

Table 1 presents the condition means for the two estimations of the percentage of hits scored by the Ss' "partners." Since reliable individual differences in the judgments made by the various Ss were revealed by analysis of variance, the "Ss within conditions" term is employed as the error term.

The major hypothesis regarding the perceived merit of the "partner's" opinions is supported by the highly significant main effect for Liking ($F = 10.98$, $p < .01$). Liked "partners" are judged to have a higher percentage of hits (correct opinions) than are less liked "partners."

The hypothesis dealing with the combination of liking and past proficiency also is upheld. Because of the absence of significant differences between blocks of series, the t tests are based upon the condition means for all eight series. The Ss in the High Like, High Prof. condition report a significantly higher percentage of correct opinions for their "partners" than the Ss in the other conditions. Similarly, Ss in the Low Like, Low Prof. condition judge their "partners" as being correct significantly less often than the Ss in either the High Like, Low Prof. or Low Like, High Prof. conditions.

Conformity

Analyses of variance of the conformity scores based upon the combined eight series did not yield any significant F ratios, perhaps be-

² This selection is justified to a certain extent because we are interested in testing the effects of differences in liking, not whether the present instructions can successfully produce differences in liking. The present screening, of course, imposes a limitation upon the population to which the results can be generalized. A possibility remains that the selected subjects would uniformly be located toward the extreme end of a personality dimension that enters into the relationship between liking, perceived opinion merit, and conformity.

TABLE 1
MEAN PERCEIVED ACCURACY OF "PARTNER"

Time of Measurement	High Proficiency		Low Proficiency	
	High Like	Low Like	High Like	Low Like
After Series 1-5:	64.5	47.0	56.9	35.8
After Series 6-8:	62.3	45.0	49.8	42.0

TABLE 2
MEAN PER CENT OF POSSIBLE DISTANCE AWAY FROM "PARTNER" IN THE FIRST SERIES (ARC SINE TRANS.)

High Proficiency		Low Proficiency	
High Like	Low Like	High Like	Low Like
32.40	48.90	42.10	45.20

cause of task withdrawal reactions in the Low Prof. condition. That the latter effect may have played a part is suggested by the fact that after the first series of targets, three of the Ss in the Low Like, Low Prof. condition had consistently written down as their decisions the exact estimates provided by either the "observer" or the "map," instead of compromising between these two sources of information as generally was the case.³

Because this type of extreme task reaction was not evident during the first series of targets, analyses of variance were conducted employing the normalized percentage data for this series. Nevertheless, the possibility remains that even at this time at least several Ss in the Low Like, Low Prof. condition were relatively indifferent to their assigned task of deciding between the two sources of information, so that even the first series data in this condition might be relatively unreliable. The condition means are given in Table 2.

The major hypothesis with regard to behavioral conformity is upheld by a significant main effect for Liking ($F = 11.11$, $p < .01$). Combining both proficiency conditions, Ss exhibit greater conformity to the opinions expressed by liked partners than to the less-liked partners' opinions.

The analysis of variance also revealed an unexpected interaction. The difference between the High and Low Like conditions is

³ It is worth noting that this presumed withdrawal from the task took place only in the Low Like, Low Prof. condition. The Ss in this condition did not have the satisfactions stemming from either liked or proficient "partners."

significant, by t test, for the High Prof. conditions ($p < .01$), but is nonsignificant for the Low Prof. conditions. Since Ss in the Low Prof., High Like condition perceive their "partners" as having significantly more correct opinions than Ss in the Low Prof., Low Like condition, it is possible that this nonsignificant difference in degree of conformity in the Low Prof. condition is an artifact of the hypothesized task withdrawal reactions.

The hypothesis dealing with the combined effects of liking for the "partner" and "partner" proficiency upon conformity to his opinions also is given some support. The Ss in the High Like, High Prof. condition have a significantly greater tendency to conform to their "partners'" opinions than the Ss in the other conditions. (One t value is significant beyond the .05 level, while the others are significant at less than the .01 level.) The other implication of this hypothesis, that Ss in the Low Like, Low Prof. condition would have the lowest level of conformity, is not supported, again perhaps as a result of unreliability of the data in the Low Like, Low Prof. condition.

DISCUSSION

Recent theoretical interpretations of social influence explain the frequently observed high degree of conformity to the opinions advocated by attractive (liked) groups by postulating either that pressure toward uniformity are relatively strong in the attractive groups (3, 5) or, to take another recent theoretical model (7), that attractive groups have relatively great "power" over the individual members. Both of these formulations appear to imply some punishment for nonconformity. In contrast, the present hypothesis is that individuals also may conform to and adopt the opinions held by like groups because liking has increased the perceived merit of the group's opinions.⁴ While the present experiment has

not demonstrated unequivocally that the effect of liking for others upon the judged merit of the others' opinions produces the subsequent conformity to their opinions, the obtained findings generally appear to be consistent with this interpretation.

That the present hypothesis can account for some of the observed conformity behavior in the studies by Festinger and his colleagues also is suggested by the nature of their experimental situations. Typically, the Ss in these experiments discussed issues about which they were probably not greatly concerned, without much time to think about the opinions they were asked to express. The Ss in these experiments may have been uncertain about the correctness of their initial opinions, and, liking the others in their groups, have perceived the others' opinions as being better. Such uncertainty about one's initial opinion should be greater among opinion deviates, as Festinger, Gerard, *et al.* (6) have shown to be the case.

The present findings also have some implications for theoretical conceptions of the correlates of group attractiveness. Back has demonstrated (1) that attractive groups are more likely than unattractive ones to influence the behavior of the individual members, regardless of the source of attraction. It is not at all certain, however, that other sources of attraction than liking for the group's members also affect the perceived merit of the group's behavior or opinions.

REFERENCES

1. BACK, K. The exertion of influence through social communication. *J. abnorm. soc. Psychol.*, 1951, 46, 9-24.
2. BERKOWITZ, L., & LEVY, B. I. "Pride in group

"strain toward symmetry" appear to prohibit the use of his model in the present situation. It is suggested that the "strain toward symmetry" arises as people learn the two advantages of symmetrical orientations: (a) ready predictability of another's behavior with regard to other Xs, and (b) validation of one's own orientation toward X. These advantages obviously do not hold in this case since the Ss know they will not directly experience the Xs themselves, and there would be little reason for the present Ss to achieve symmetry if these were the only benefits.

To tentatively advance another explanation for the present results, an individual who likes another associates this other with his self. Therefore, to the extent that he has a need to enhance the merit or value of his own self, he also will strive to enhance the merit of the liked person.

⁴ This hypothesis is similar to (but not identical with) that advanced by Newcomb (11) in his extension of Heider's general principle of "balanced states." According to Newcomb, S's "strain toward symmetry" in a matrix consisting of himself, his orientation (attitude) toward B (S's partner in the present sense), and his orientation toward X (an opinion advanced by B) increases with the valence of the objects of co-orientation. Therefore, the greater S's attraction toward B, the more likely S's orientation toward X will be "symmetrical" with his orientation toward B. However, Newcomb's hypotheses concerning the origin of this

- performance" and group task motivation. *J. abnorm. soc. Psychol.*, 1956, **53**, 300-306.
3. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271-282.
4. FESTINGER, L. An analysis of compliant behavior. In M. Sherif (Ed.), *Group relations at the cross-roads*. New York: Harper, 1953.
5. FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117-140.
6. FESTINGER, L., GERARD, H., HYMOVITCH, B., KELLEY, H., & RAVEN, B. The influence process in the presence of extreme deviates. *Hum. Relat.*, 1952, **5**, 327-346.
7. FRENCH, J. R. P., JR. A formal theory of social power. *Psychol. Rev.*, 1956, **63**, 181-194.
8. GILCHRIST, J. C. The formation of social groups under conditions of success and failure. *J. abnorm. soc. Psychol.*, 1952, **47**, 174-187.
9. KIDD, J., & CAMPBELL, D. The conformity to groups as a function of group success. *J. abnorm. soc. Psychol.*, 1955, **51**, 390-393.
10. MAUSNER, B. The effect of one partner's success or failure in a relevant task on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 557-560.
11. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393-404.

Received July 2, 1956.

STUDIES IN THE PHENOMENOLOGY OF EMOTIONS¹

JACK BLOCK

Institute of Personality Assessment and Research, University of California

THE present paper describes an application of Osgood's "semantic differential" method (6), to the study of the phenomenology of emotions. Data dealing with sex and cultural differences are reported.

The semantic differential, as already attested by its widespread use in a variety of research enterprises, is a most versatile technique. By a standardized connotative description of various objects or concepts, the meaning of meaning is given what seems to be quite an appropriate quantification, making fruitful analyses possible.

The feature of interest here is that the method, when applied to the description of emotions, appears to provide operational access to the phenomenology of emotions. Phenomenologists, as for example Sartre in his small volume on emotions (10), have been persistent, and in the writer's opinion justified, in criticizing psychologists for neglecting the study of feelings. Although the criticisms have frequently stemmed from basically antiscientific philosophical positions, it does seem that psychology should be more concerned with the nature or meaning or conditions of private emotional experience. Some preliminary work suggested that the semantic differential could quantify affective qualities in an unstereotyped way and the studies reported here were therefore planned.

A word of caution first. There is always a degree of arbitrariness in positing an operation as a relevant or sufficient manifestation of a construct. What is required is that the arbitrariness be of a "reasonable" kind, where "reasonableness" unfortunately can be only vaguely evaluated. In suggesting that the procedure for description of emotions by the semantic differential could be taken as an operational index of the actual quality of an

individual's emotional experience, two considerations were persuasive.

1. There is no dictionary definition of affects such as *humiliation*, *grief*, *boredom*, and so on, in connotative terms. Moreover, there is little, if any, institutional or cultural referent to help in the connotative description of emotions. We have not learned simple conventions to employ when the rather unusual task of describing emotions is presented to us. Indeed, as will be seen in Table 1 below, *envy* in actuality is described as somewhat more *red* than *green*, as the verbal cliché would dictate. How then is the task coped with by the individual? It would seem that in order to fulfill the demands of the test situation, the *S* can only turn inward for the basis of his report and what is inward is by implication his private emotional experience.

2. The more general properties of language habit systems, to the extent they do operate in the description of emotions, seem to reflect experiential qualities rather than initially fortuitous and phenomenologically irrelevant response tendencies. See, for example, the work on synesthesia (3, 9, 12) and physiognomic perception (12).

The proposed correspondence of construct and measure, of course, must receive its support from the network of results it engenders. The two studies reported here bear on the proffered interpretation, and it is to these we now turn. The specific expectancies underlying the present researches were:

a. As measured by the semantic differential, the phenomenology and interrelations of emotions in men are different from the phenomenology and interrelations of emotions in women.

b. Across cultures, differences in the phenomenology and interrelations of emotions are related to independently evaluated differences in cultural values and emphases.

STUDY I

The form of the semantic differential employed in the first study followed closely that described by Osgood (6). Fifteen emotions, selected *a priori* but with the intention of com-

¹ This investigation, part of a larger research project, was supported by research grant M-1078 from the National Institute of Mental Health of the United States Public Health Service. Mrs. Anne von der Lippe first recognized the interest of a cross-cultural comparison of phenomenologies. The translation of the description of emotions procedure, its administration in Norway, and many of the computations are her work. I am privileged to present her data here.

TABLE 1

RANK ORDERING OF CONNOTATIVE DIMENSIONS IN THE DESCRIPTION OF EMOTIONS (FEMALE SAMPLE)

Connotative Dimension	Emotion														
	Love	Anger	Envy	Humiliation	Nostalgia	Pride	Fear	Boredom	Contentment	Sympathy	Grief	Worry	Elation	Guilt	Anticipation
HIGH-low	4*	4	7	12.5	14	1	6	19.5	3	8	16	14	1	16	1
GREEN-red	11	20	13	14.5	3	13	15	4	8	5	5	10	11	12	13
WEAK-strong	20	18	14	7	6	20	10	3	13	11	5	7	17	4.5	19
ROUGH-smooth	12	3	6	10	18	10	4	12	15.5	20	9	3	12	4.5	10
ACTIVE-passive	5	1	4.5	11	20	2	5	18	10	13.5	14	5	2	15	2
EMPTY-full	19	13	9	4	7	19	11	1	17	18	7	9	19	9	20
SMALL-large	18	15	11	5	13	17	16	7	14	17	14	13	18	8	16
COLD-hot	15	16	10	14.5	8	15.5	8	5	12	16	3	8	14	7	15
CLEAR-hazy	9	6	12	17	19	3	14	15	6	9	14	15	5	14	8
YOUNG-old	8	8	8	6	16	4	9	11	9	12	18	16	7	13	7
GOOD-bad	1.5	12	20	19	10	6	17	17	2	1	19	20	3	19	4.5
PEACEFUL-ferocious	6	19	18	16	2	11	20	9	1	2	12	17	8	17	11
SICK-healthy	17	10	4.5	3	12	15.5	7	6	18.5	15	5	4	16	3	17
ANGULAR-rounded	14	5	2	8	17	8	3	13	15.5	19	11	6	13	6	9
TENSE-relaxed	13	2	1	1	15	5	1	10	18.5	13.5	2	1	10	2	4.5
SAD-happy	16	7	3	2	4	14	2	2	20	7	1	2	20	1	18
SOFT-loud	7	17	16	9	1	18	18	8	7	3	8	12	15	11	14
WET-dry	10	11	15	12.5	11	12	12.5	14	11	10	10	11	9	10	12
BEAUTIFUL-ugly	3	14	19	20	5	9	19	16	4	4	20	19	6	20	6
FRESH-stale	1.5	9	17	18	9	7	12.5	19.5	5	6	17	18	4	18	3

* Low ranks mean that the capitalized end of the connotative dimension applies; high ranks, that the small-lettered pole is indicated.

prehensive coverage of the affective sphere, were used as the concepts to be defined. Each concept was evaluated in terms of twenty connotative 7-point scales. The emotions selected and the defining scales used may be read from Table 1.

The semantic differential form was administered to 40 male and 48 female college students in an undergraduate psychology course. The Ss described the emotions by marking appropriate points along each defining scale. The S's orientation was specifically to describe each affect as personally experienced, not as perceived in others or as cultural referents would suggest.

In order to compare the two groups, mean descriptions of each emotion were compiled for each sex separately. For a given emotion, the scores assigned to the twenty scales first were summed across the 40 males and, separately, across the 48 females. Then, for each emotion, the twenty sums for men, and the twenty sums for women were separately rank-ordered, to provide a quick comparability between descriptions.²

² Ss, as seems introspectively appropriate, tended to emphasize middle intervals on certain affects such as *nostalgia*. Although the rank orderings of these affects are therefore of lower reliability because based on scores with low variability, no bias is therefore introduced into the correlations between affects.

In Table 1 are listed the rank orderings of the connotative scales for each of the fifteen affects, for the female sample. These data are presented here for their interest as a portrayal of the phenomenological qualities of the emotions considered. The data for the male sample are not presented because with the exception of the emotion of *grief*, the connotative descriptions of emotions by the men and by the women correlated very highly (see Table 4), the range being from .84 to .98, with a mean, computed via the *z* transformation, at .94. The rank-order correlation between the sexes for *grief* was only .66. What the meaning of this discrepancy may be is obscure, but for women *grief* is described as relatively *low*, *green*, *weak*, *smooth*, and *rounded*, while for men *grief* is relatively *high*, *red*, *strong*, *rough*, and *angular*.

The intercorrelations between the emotions were computed separately for both men and women. The matrix based upon the female sample is presented in Table 2. Because of the high correlations already noted between the male and female descriptions of emotions, the correlation matrix based upon the male sample is necessarily similar to that for the females. Enough latitude remains, however, so that although pair-wise, male and female descriptions of emotions are similar, the or-

TABLE 2
INTERCORRELATION OF EMOTIONS IN A FEMALE SAMPLE

Emotion	Emotion														
	Love	Anger	Envy	Humiliation	Nostalgia	Pride	Fear	Boredom	Contentment	Sympathy	Grief	Worry	Elation	Guilt	Anticipation
Love	—														
Anger	+.15	—													
Envy	-.48	+.68	—												
Humiliation	-.73	+.29	+.74	—											
Nostalgia	-.03	-.81	-.55	-.14	—										
Pride	+.68	+.62	+.17	-.36	-.63	—									
Fear	-.37	+.74	+.91	+.62	-.29	+.31	—								
Boredom	-.82	-.51	+.19	+.59	+.50	-.82	+.06	—							
Contentment	+.83	-.28	-.75	-.85	+.20	+.41	-.69	-.59	—						
Sympathy	+.66	-.40	-.70	-.57	+.59	+.15	-.63	-.24	+.73	—					
Grief	-.73	-.12	+.47	+.60	+.28	-.56	+.44	+.77	-.71	-.31	—				
Worry	-.67	+.38	+.81	+.73	-.22	-.24	+.80	+.47	-.87	-.62	+.81	—			
Elation	+.88	+.36	-.27	-.68	-.37	+.87	-.16	-.90	+.74	+.42	-.71	-.54	—		
Guilt	-.85	+.16	+.69	+.82	-.09	-.47	+.66	+.66	-.92	-.63	+.84	+.88	-.78	—	
Anticipation	+.82	+.56	-.04	-.48	-.48	+.92	+.07	-.91	+.55	+.26	-.65	+.36	+.92	-.61	—

ganization or interrelation of emotions within each sex may be quite different. For this reason, the two correlation *matrices* were correlated with each other.

A fundamental test of the functional equivalence of two sets of data is that there be an isomorphism of the relationships among the variables constituting each data set. That is, if there exists a strong relationship between emotions A and B in the one data set, there should also exist a strong relationship between these same variables in the second data set. The same pattern of covariation among variables must be present in the two corresponding matrices. An over-all method for characterizing the degree of isomorphism is simply to treat the correlations between variables as *scores* and to correlate all of these scores paired across data sets. The coefficient resulting from this inter-matrix correlation is thus an index of the similarity of covariation within the two sets of data.

The intercorrelation of the two matrices, computed over the 105 separate cells, was .93. A coefficient of this magnitude means that, as measured by the semantic differential, emotions interrelate in the two sexes with full equivalence. This finding is contradictory to initial expectation and suggests that to the extent the description of emotions is a measure of the quality of affective experience, the differences in expressed emotionality between men and women are not due to differences at the level of experience. Rather, these overt

differences may be due to the learning of different behavioral patterns for the expression of emotions or to different sensitivities to particular presses.

A centroid method factor analysis was then carried out, starting from the matrix shown in Table 2, in order to dimensionalize the interrelationships among the emotions. Three factors appear to exhaust the meaningful variance. The original and rotated factor matrices are to be seen in Table 3. Factor rotation was accomplished by means of the analytic "quartimax method" (5), with a final 45° rotation of Factors I and II in order to increase the psychological meaning of the factors.

For two reasons, the results of the present factor analysis may be accepted as more reproducible than is usually the case. First, the fully equivalent correlation matrix based upon the male sample indicates that the identical unrotated factors would emerge in a second analysis. Second, the objective rotational method employed eliminates much of the ambiguity that otherwise complicates the interpretation of factor analytic results.

Factor I is defined at one end of the continuum by *contentment*, *love*, *elation*, *sympathy*, and *anticipation*, all of these being *joie de vivre* emotions. At the negative end of the dimension are found *guilt*, *worry*, *humiliation*, *grief*, *envy*, *fear*, and *boredom*, all of these being less happy conditions of man. The factor dimension involved here is clearly the *pleasantness-unpleasantness* polarity, the hedonic orientation

TABLE 3
ORIGINAL AND ROTATED FACTOR MATRICES
(All decimals omitted)

Emotion	Original Factor Matrix				Rotated Factor Matrix			
	I ^(a)	II ^(a)	III ^(a)	<i>h</i> ²	I ^(a)	II ^(a)	III ^(a)	<i>h</i> ²
Love	-888	344	203	948	862	396	221	949
Anger	215	912	065	882	-223	901	-143	882
Envy	772	583	144	957	-785	581	-059	957
Humiliation	872	060	062	768	-874	054	-042	769
Nostalgia	-136	-806	370	805	096	-712	538	806
Pride	-450	818	175	902	429	847	050	904
Fear	665	605	394	963	-703	657	192	963
Boredom	667	-688	113	931	-675	-665	181	931
Contentment	-967	-070	056	943	956	-036	170	944
Sympathy	-736	-350	457	873	684	-232	593	873
Grief	774	-336	390	864	-809	-267	368	861
Worry	894	188	255	900	-917	215	117	901
Elation	-776	587	102	957	760	612	060	956
Guilt	961	-071	108	940	-966	-070	020	939
Anticipation	-610	752	148	960	590	780	055	960

so fundamental for the individual's adaptation and behavior.

Factor II is defined by *anger*, *pride*, *anticipation*, *fear*, *elation*, and *envy* at one end of the continuum and by *nostalgia* and *boredom* at the other end. This dimension appears to reflect the extent to which the individual is organismically involved in the affect. It is not affective tone, as in Factor I, but affective intensity or *level of activation* (4) which is represented here.

Factor III reveals a most interesting and less anticipated ordering of the emotions. It is defined by *sympathy*, *nostalgia*, *grief*, and, to a lesser extent, *love*. The factor is unipolar and appears to encompass those emotions issuing from *interpersonal relatedness* and considerations of impermanence.

The first two of these factors are rather obvious and seem to correspond with those suggested by Schlosberg (11) in his analyses of emotion. The third factor cannot be linked to previous findings.

The present usage of the semantic differential differs from that of Osgood. Where Osgood correlates scales (connotative dimensions), the present approach correlates concepts (emotions). The difference in emphasis is analogous to that involved in the *R* versus *Q* distinction (2). The alternate approaches may lead to different results in the case where the pattern of covariation among scales is a function of the concepts employed. Osgood has already reported evidence for such concept-scale interaction (7).

It is perhaps for this last reason that the factor-analytic results of Study I can be related only in part to those reported by Osgood and Suci (8). The first two factors found in the present research can be linked roughly with the *evaluation* and *activity* factors of Osgood and Suci. But the *interpersonal relatedness* factor found here does not seem reconcilable with their *potency* factor or any of the residual factors they report.

STUDY II

The second study involved cross-cultural comparisons in the description of emotions. The description of emotions procedure was translated into Norwegian (with the exception of the *soft-loud* scale) and administered to 34 individuals in Norway.³ Twenty of the group were male and fourteen were female. Because of the finding of no sex differences in the first study and because of the otherwise rather small *N*s available for analysis, these two groups were combined. There is, of course, a certain risk entailed by the cross-cultural extrapolation. Some age and socioeconomic differences between the Norwegian and American samples also exist in that the Nor-

³ A copy of the Norwegian translation has been deposited with the American Documentation Institute. Order Document No. 5135 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.25 for 35 mm. microfilm or \$1.25 for 6 by 8 inch photocopies. Make check payable to Chief, Photoduplication Service, Library of Congress.

TABLE 4

CORRELATIONS BETWEEN DESCRIPTIONS OF EMOTIONS
BY NORWEGIANS AND AMERICANS

Emotion	Norwegian-American Correlation	American male-American female Correlation
Love	.90	.95
Anger	.89	.97
Envy	.76	.95
Humiliation	.57	.84
Nostalgia	-.07	.86
Pride	.93	.88
Fear	.68	.95
Boredom	.81	.94
Contentment	.93	.98
Sympathy	.80	.92
Grief	.23	.66
Worry	.81	.95
Elation	.67	.95
Guilt	.88	.94
Anticipation	.52	.96

TABLE 5

THE TWENTY LARGEST CROSS-CULTURAL DIFFERENCES
IN THE INTERRELATION OF EMOTIONS

Emotion	Inter- r among Norwegians	Inter- r among Americans
Grief-Guilt	.06	.83
Sympathy-Love	.88	.58
Worry-Grief	.20	.84
Nostalgia-Anger	.43	-.71
Elation-Love	.58	.88
Envy-Love	-.73	-.40
Pride-Love	.85	.57
Nostalgia-Love	.77	.08
Sympathy-Anger	.31	-.47
Elation-Anticipation	.35	.91
Anticipation-Pride	.31	.88
Pride-Envy	-.44	.36
Nostalgia-Pride	.55	-.61
Fear-Anger	.19	.79
Sympathy-Pride	.67	-.11
Sympathy-Guilt	-.86	-.59
Sympathy-Contentment	.96	.74
Boredom-Elation	-.50	-.91
Grief-Envy	-.29	.59
Grief-Contentment	-.03	-.79

wegian sample is slightly older over-all, and with greater age range, and averages somewhat lower in social class. To the extent these several differences are relevant, the comparisons across cultures should show lack of correspondence, and interpretation of obtained differences becomes more difficult.

The data from the Norwegian sample were treated as were the American data. For each emotion, the correlation between the description by Norwegians and the description by

American males was then computed. These may be read from Table 4. With the exception of *grief* and *nostalgia*, the correlations are reasonably high.

The intercorrelation matrix of the 15 emotions was derived, and this matrix in turn correlated with the matrix for the American males, in order to evaluate the extent of equivalence in the way emotions are *organized* in the two cultures. The inter-matrix correlation coefficient was .69 (.70 when three emotions were deleted for which there were *a priori* reservations as to the adequacy of the translation). A correlation of this magnitude is indicative of an appreciable degree of congruence between the two groups, particularly in the light of the potentially attenuating factors previously mentioned. As measured by the semantic differential, it would appear that emotions stand in roughly the same relation to each other in the American and Norwegian cultures. What the interrelations may be in other cultures remains to be assessed, but it does not seem an unlikely hypothesis that the universal experiences of man generate a common phenomenological background. Some additional evidence for this interpretation is to be found in a recent study by Asch (1).

Although the two cultures correspond greatly in the way their emotions are organized, the original interest was in the nature and meaningfulness of the differences observed. To pursue this concern, the twenty greatest differences in the way emotions were related in the two cultures were selected. That is, for a given pair of emotions, e.g. *grief-guilt*, the correlation in one culture was .83 and in the other culture .06. Differences approaching this magnitude account for the lack of total congruence between the two matrices.

The twenty pairs of emotions chosen are listed in Table 5. If the differences between cultures are meaningful, it was hypothesized, judges with some knowledge of both cultures should be able to predict the nature and direction of the differences found. Sixteen judges, psychologists and psychiatrists, attempted the prediction task, with the alternatives randomly arranged. Where chance would offer only 10 correct choices, the judges' consensus was accurate in 17 of the 20 estimates, an occurrence significant beyond the .01 level. For the judges individually, ac-

curacy scores ranged from 8 to 16, with a mean of 13.8. The differences between the two groups, despite the strong congruence in the way they depict emotions, do appear to be meaningful when evaluated in this fashion.

What these differences suggest, so far as cultural emphases and values are concerned, is open to varying interpretation. The adequacy of a crosscultural translation can perhaps always be questioned. There is the possibility, slight in the writer's estimation, that the differences observed reflect class distinctions rather than cultural differences. And certainly a number of more psychological hypotheses present themselves. For example, the strong relationship between *grief* and *guilt* in the American culture where they are orthogonal among Norwegians has psychoanalytic import and suggests a confounding of the two emotions by Americans. But it is probably wisest at the present stage of information to delay interpretation until further research evidence is in.

SUMMARY

Two studies are reported employing the semantic differential as a means of quantifying the phenomenology of emotions. In the first study, American men and women college students were found to describe emotions connotatively in a highly similar way. Factor analysis of these data revealed three underlying dimensions: a *pleasantness-unpleasantness* factor, a *level of activation* factor, and a factor derived from emotions involving *interpersonal relatedness*.

The second study compared a sample of

Norwegians with Americans with respect to their descriptions of emotions. Although an appreciable congruence was observed, the differences in phenomenological organization proved to be predictable and hence presumably reflective of differences in cultural values.

REFERENCES

1. ASCH, S. On the use of metaphor in the description of persons. In H. Werner (Ed.), *On expressive language*. Worcester, Mass.: Clark Univer. Press, 1955. Pp. 29-38.
2. BLOCK, J. The difference between *Q* and *R*. *Psychol. Rev.*, 1955, **62**, 356-358.
3. KARWOSKI, T. F., ODBERT, H. S., & OSGOOD, C. E. Studies in synesthetic thinking: II. The role of form in visual responses to music. *J. gen. Psychol.*, 1942, **26**, 199-222.
4. LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951, Pp. 473-516.
5. NEUHAUS, J. O., & WRIGLEY, C. The quartimax method: an analytic approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, **7**, 81-92.
6. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197-237.
7. OSGOOD, C. E. Report on development and application of the semantic differential. Unpublished manuscript, Inst. of Communications Research & Dept. of Psychol., Univer. of Illinois.
8. OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325-338.
9. RIGGS, L. A., & KARWOSKI, T. Synaesthesia. *Brit. J. Psychol.*, 1934, **25**, 29-41.
10. SARTRE, J.-P. *The emotions: outline of a theory*. New York: Philosophical Library, 1948.
11. SCHLOSBERG, H. Three dimensions of emotions. *Psychol. Rev.*, 1954, **61**, 81-88.
12. WERNER, H. *Comparative psychology of mental development* (Rev. ed.), Chicago: Follet, 1948.

Received July 3, 1956.

EFFECTS OF BIOLOGICAL AND SOCIAL MOTIVATION ON SCHIZOPHRENIC REACTION TIME

GERALD ROSENBAUM, WILLIAM R. MACKAVEY, AND JAMES L. GRISELL

Wayne State University and Northville State Hospital

THEORY and research on the nature of schizophrenia appear to have been converging in recent years toward an attempt to derive disturbances in schizophrenic performance from a central motivational pathology. Hoskins (9), Babcock (3), and others (1, 15) have accumulated evidence pointing to a deficit in the basic biological motivation and the innate equipment necessary for coping with environmental stresses. On the other hand, Sullivan (19), Cameron (6), and Hunt and Cofer (10), among others (14, 17, 21), have marshaled support for the intactness of the schizophrenic's capacity for biological arousal. They have suggested that characteristic disturbances in schizophrenic language, thought, attention, and affect may be most accurately regarded as extensions of an acquired deficit in the social drives, intensified during the process of social adaptation.

A recent study by Cohen (7) was designed to investigate this issue by comparing schizophrenic performance under biological motivation with performance under ordinary social incentives. Finding that shock induced motivation improved motor learning more in schizophrenics than in normals, Cohen suggested that primary motivation and reward may be employed to compensate for inadequate social motivation in effecting learning in schizophrenics. The present experiment was designed to test the related hypothesis that characteristic disturbances in schizophrenic performance can be eliminated under biological shock motivation.

A task which has consistently yielded decrements in schizophrenic performance is the reaction time procedure originally employed by Huston, Shakow, and Riggs (12) and later developed into a schizophrenic set index by Rodnick and Shakow (18). Several repetitions of the original study under slightly different conditions (8, 11, 13, 20) have confirmed the essential findings that schizophrenics are slower, more variable, and have more difficulty in maintaining a regular preparatory set than normals. If such disturbances in schizophrenic

attention and performance can be accounted for by impaired social motivation, then the arousal of a biological drive such as pain should improve schizophrenic performance to a normal level.

METHOD

To test this hypothesis, schizophrenic and normal Ss were tested on a reaction time task, in which a buzzer signal was presented at several regular and irregular preparatory intervals. Half the Ss in each group were then immediately retested on the same task under biological motivation, activated by a shock emitted by the response key at the same instant that the buzzer signal occurred. The other half of each group was retested as a control for the effects of practice without shock. It was predicted that schizophrenics would show significantly more improvement than normals under shock conditions, thereby reaching the same level as the normals in reaction time, variability, and ability to maintain a regular set.

Subjects

A group of 32 schizophrenic patients and a group of 32 employees from Northville State Hospital, matched for age and sex, were each divided into an experimental and a control group. Each of these four groups contained 10 males and 6 females. The mean ages in years for the four groups were as follows: schizophrenic-experimental 33.9, range 17-52; schizophrenic-control 32.3, range 24-47; normal experimental 34.9, range 24-52; and normal-control 33.9, range 24-52. All patients had been hospitalized less than three years.

Apparatus

The stimulus presentation and recording units of the reaction time apparatus were housed in a portable wooden cabinet 20.5 in. long, 17 in. wide, and 10.5 in. high. The S's response unit consisted of a telegraph key and copper plate mounted on the far end of a 22 x 17 in. flat panel, hinged to the cabinet. The illumination of a 25-w. bulb, centered on the front face of the cabinet 8 in. above the panel, served as the ready signal.

The ready signal and a Hunter Decade Interval Timer were simultaneously activated by E's closing a push-type switch. Depression of the telegraph key by S in response to the ready signal closed a holding relay, permitting E to release his switch. At the termination of the interval previously set on the interval timer, a door bell buzzer and Standard Electric Timer were instantaneously activated. Release of the response key to the buzzer signal broke the common circuit to the

TABLE 1
REACTION TIME SCORES IN .01 SEC. OF SCHIZOPHRENICS AND NORMALS

	Schizophrenics				Normals			
	Experimental		Control		Experimental		Control	
	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)
Initial Test								
Mean	35	39	49	79	23	25	24	24
Median	32	37	31	53	23	25	24	22
Retest								
Mean	19	33	37	93	17	19	22	24
Median	17	29	31	40	17	20	22	22

TABLE 2
VARIABILITY SCORES OF SCHIZOPHRENICS AND NORMALS

	Schizophrenics				Normals			
	Experimental		Control		Experimental		Control	
	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 10)
Initial Test								
Mean	158	162	243	584	69	95	66	70
Median	97	153	119	346	61	94	66	64
Retest								
Mean	64	137	109	467	46	60	49	54
Median	43	83	91	211	44	54	42	55

buzzer and clock. Reaction times were read directly to the nearest .01 sec.

The shocks, generated by a transformed 50-v. current in parallel with the buzzer and clock, were administered through a copper electrode on the surface of the response key and the copper plate on which *S* rested his hand. The setting of a toggle switch determined whether or not shock would accompany the buzzer.

Procedure

All *Ss* were initially tested under the social instructions that they were taking a reaction time test requiring them to depress the key to the light signal and to release it as rapidly as possible at the sound of the buzzer. Each *S* was then given 10 practice trials at each of the three preparatory intervals, 2, 5, and 10 sec., first presented in that order in regular homogeneous blocks and then in an irregular randomized order. The first two reactions at each of the three regular intervals and the first six reactions of the irregular procedure were counted as practice trials and excluded from all computations, leaving 48 recorded test trials. Following this initial run of 60 trials, a 5-min. rest was introduced to reduce fatigue effects. The control group of schizophrenics and the comparable normals were then retested under the identical conditions of ordinary social motivation. Both experimental groups were also retested on the same procedure but were instructed that they would "... feel some skin stimulation on the key at the same moment the buzzer

sounds." Shocks were administered to the experimental groups on all 60 retest trials. All anticipatory responses were immediately repeated, so that every *S* was required to react to the buzzer on 120 trials.

RESULTS

The major results of the experiment are presented in Tables 1 and 2, which show the reaction time and variability scores of the schizophrenic and normal groups, broken down for sex, on both the initial test and retest. The individual reaction time scores on which the group data in Table 1 are based were obtained for each *S* by computing the overall mean of his 48 response latencies to the three test intervals in the regular and irregular orders. The variability scores comprising the basic data for Table 2 were computed for each *S* by obtaining the range of his eight reaction time scores for each test interval in the regular and irregular procedures, and then summing these six ranges. In view of the marked variability of the schizophrenic groups, medians as well as means are shown in the tables to permit an evaluation of the group trends relatively independently of the extreme scores.

Social Motivation

Support for the hypothesis that the schizophrenics showed an initial deficit in social motivation is afforded by examination of the first two rows of Tables 1 and 2, which show that all four schizophrenic groups were consistently slower and more variable than the comparable normal groups. To evaluate these trends statistically, all scores were transformed to the reciprocals of their square roots so as to eliminate the heterogeneity of variance occasioned largely by the extreme variability of the schizophrenic groups. Analysis of variance of the transformed reaction time scores for the initial test showed that the schizophrenics were significantly slower than the normals (.01 level), and that females were significantly slower than males (.05 level). No significant differences were obtained between experi-

mental and control groups, nor were any of the interactions between the three main sources of variance significant. This finding indicates that there was no appreciable selection bias in assigning Ss to experimental and control groups. Similar analysis of the initial test variability scores showed that the diagnostic category was the only significant source of variance (.01 level).

Biological Motivation

The differential effects of biological and social motivation on the schizophrenic and normal groups are shown in the retest data of Tables 1 and 2. Inspection of these group trends reveals that the experimental groups receiving shock tended to be faster and less variable than the control groups, and that this tendency was more marked in the schizophrenic groups. Transformation and subsequent analyses of variance of these scores showed that the normals, experimental groups, and males were all faster and less variable than their comparable groups at the .01 level of significance. The results of the analysis of variance of the variability scores are presented in Table 3.

Since both schizophrenics and normals showed improvement in performance with shock, the interaction between these two main effects constitutes a test of the hypothesis that the experimental groups of schizophrenics improved over their comparable controls more than did the normals. Table 3, and careful scrutiny of the retest data in Table 2, reveal that shock did significantly reduce variability in both male and female schizophrenics, but produced only a slight decrease in the variability of normals. While this same trend seems to appear in the reaction time data in Table 1, it appears less marked, and statistical analysis revealed no significant interaction. None of the remaining interactions with the sex factor proved to be significant for either the reaction time or variability data.

Preparatory Set Data

The principal findings with regard to the effects of regular and irregular preparatory sets are presented in Table 4. While other investigators (11, 12, 13, 18, 20) have consistently found that normals react more rapidly with the regular than with the irregular procedure,

TABLE 3
SUMMARY OF ANALYSIS OF VARIANCE OF TRANSFORMED
VARIABILITY SCORES ON RETEST

Source	df	Mean Square	F
Diagnosis (D)	1	.0713	9.90**
Motivational Condition (M)	1	.0539	7.49**
Sex (S)	1	.0439	6.09**
D × M	1	.0319	4.43*
D × S	1	.0062	—
M × S	1	.0042	—
D × M × S	1	.0001	—
Within Groups	56	.0072	—
Total	63		

* Significant beyond the .05 level.
** Significant beyond the .01 level.

TABLE 4
MEAN REACTION TIMES IN .01 SEC. AT DIFFERENT
PREPARATORY INTERVALS

Experimental Group	Initial Test			Retest		
	Interval in Sec.			Interval in Sec.		
	2	5	10	2	5	10
Schizophrenic Exp.						
Regular	38	34	32	21	26	23
Irregular	45	38	32	28	24	24
Schizophrenic Control						
Regular	57	64	70	61	51	58
Irregular	67	51	52	60	63	59
Normal Exp.						
Regular	24	23	24	16	17	18
Irregular	28	23	23	21	19	18
Normal Control						
Regular	25	23	23	21	21	22
Irregular	27	23	22	26	23	23

no reliable difference between these two procedures is apparent in either of the normal groups. Since the normal Ss were not appreciably faster on the regular than the irregular procedure, the effect of biological and social motivation on the typical schizophrenic failure to benefit from the regular preparatory set could not be investigated.

Further inspection of Table 4 indicates that only one preparatory set phenomenon was observed consistent with previous findings: the longer latencies on the 2-sec. irregular interval. These slow reactions appear to result from the inability of Ss to anticipate the trials requiring a brief response set. The present failure to replicate previous observations on preparatory set appears to be an artifact of procedural modifications, introduced because the present retest addendum made it desirable to shorten and to simplify the testing procedure. These modifications included fewer test trials at each interval, administration of the regular procedure first rather than last, and presentation of the regular intervals in the order 2, 5, 10 rather than 5, 2, 10. These modifications, incidentally, should tend to reduce the facilitating effects of warm up and practice on the temporal conditioning phenomenon inherent in the regular procedure.

DISCUSSION

The present results confirm previous findings that schizophrenics are slower and more variable than normals when performing a reaction time task under ordinary social motivation. Confirmation of previously reported sex differences favoring males (8) and a finding that females have higher variability scores than males were also obtained.

The major hypothesis, that schizophrenics would improve more than normals with shock motivation, was clearly supported for the variability scores but not for average reaction time. Evidence that schizophrenics can achieve a normal level of functioning when biologically motivated, however, can be seen in the data for males in Tables 1 and 2. The median variability and reaction time scores of schizophrenic males in the experimental group are clearly similar to those of normal males under shock motivation. These results are consistent with Cohen's findings for motor learning (7) and those of Paintal (16) for the GSR, supporting the view that the biological motiva-

tional systems of schizophrenics are unimpaired and when activated may greatly reduce the characteristic deficit in schizophrenic performance.

The very marked individual differences exhibited by schizophrenic Ss suggest, however, that this generalization does not hold for all schizophrenics. While the group trends were in the appropriate direction, two of the schizophrenic males and three of the schizophrenic females displayed mean reaction times under shock motivation that were slower than those of all comparable normals under ordinary social motivation. Furthermore, one of these schizophrenic females showed an *increase* in mean reaction time from .36 seconds without shock to .58 seconds with shock.

Clinical evaluation of these atypical individuals in the present schizophrenic sample indicated chronicity of their symptom patterns. This observation suggests that defective biological motivation may be characteristic either of the more chronic forms of schizophrenia or of some late stage of the schizophrenic process (2). This suggestion is consonant with Bellak's proposal (4) that schizophrenia may be a reaction type associated with varying biological and sociopsychological predispositions, as well as with Brackbill's recent suggestion (5) that some kind of structural brain damage may be found in those schizophrenic patients who show the classical signs of formal thinking disorder. Investigation of the relationship of the presently obtained index of defective biological motivation to variables such as age, length of hospitalization, duration of illness, presence of thinking disorder, and extent of deterioration, should facilitate an improved conceptualization of schizophrenia and the relevant dimensions of motivation.

SUMMARY

The present experiment was designed to test the hypothesis that characteristic disturbances in schizophrenic reaction time are the result of defective social motivation, and can be eliminated under biological shock motivation.

Thirty-two schizophrenic patients and 32 hospital employees, matched for age and sex, were first tested on a reaction time task in which they were instructed to lift their finger from a telegraph key in response to a buzzer presented 10 times at each of three regular and three irregular preparatory intervals. Half the

patients and normals were then immediately retested on the same task under biological shock motivation, and the other half of each group was retested as a control for the effects of practice without shock.

The results showed that the normals and the males were faster and less individually variable than the schizophrenics and the females under all conditions. Shock was also significant as a main source of variance, resulting in greater improvements in speed and variability than did practice. No difference between the regular and irregular orders was obtained for any of the groups or subgroups.

The major hypothesis was partially supported in that schizophrenics showed significantly greater improvement with shock than did normals, and shock motivation raised the median scores of the males to the level of the normals. However, the female schizophrenics and some individual males did not improve significantly with shock.

It was concluded that disturbances in schizophrenic reaction time are only partially accounted for by defective social motivation and it is suggested that a biological deficit may be involved in the reactions of more chronic schizophrenics.

REFERENCES

1. ANGYAL, A., FREEMAN, H., & HOSKINS, R. G. Physiological aspects of schizophrenic withdrawal. *Arch. neurol. Psychiat.*, 1940, 44, 621-626.
2. ARIETI, S. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955.
3. BABCOCK, HARRIET. *Dementia praecox: a psychological study*. Lancaster, Pa.: Science Press, 1933.
4. BELLAK, L. *Dementia praecox*. New York: Grune & Stratton, 1948.
5. BRACKBILL, G. A. Brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, 53, 210-226.
6. CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944, Pp. 861-921.
7. COHEN, B. D. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, 52, 186-190.
8. HALL, K. R. L., & STRIDE, E. Some factors affecting reaction times to auditory stimuli in mental patients. *J. ment. Sci.*, 1954, 100, 462-477.
9. HOSKINS, R. G. *The biology of schizophrenia*. New York: Norton, 1946.
10. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944, Pp. 971-1032.
11. HUSTON, P. E., & SENF, RITA. Psychopathology of schizophrenia and depression: I. Effect of amytal and amphetamine sulfate on level and maintenance of attention. *Amer. J. Psychiat.*, 1952, 109, 131-138.
12. HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: Reaction time. *J. gen. Psychol.*, 1937, 16, 39-82.
13. HUSTON, P. E., & SINGER, MARY M. Effect of sodium amytal and amphetamine sulfate on mental set in schizophrenia. *Arch. neurol. Psychiat.*, 1945, 53, 365-369.
14. MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenia. *J. Pers.*, 1951, 19, 359-375.
15. MAY, P. R. A. Pupillary abnormalities in schizophrenia and during muscular effort. *J. ment. Sci.*, 1948, 94, 89-98.
16. PAINTAL, A. S. A comparison of the galvanic skin responses of normals and psychotics. *J. exp. Psychol.*, 1951, 41, 425-428.
17. RICKERS-OVSIANKINA, MARIA. Studies on the personality structure of schizophrenic individuals: Reactions to interrupted tasks. *J. gen. Psychol.*, 1937, 16, 179-196.
18. RODNICK, E. H., & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, 97, 214-225.
19. SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William A. White Foundation, 1947.
20. TIZARD, J., & VENABLES, P. H. Reaction time responses by schizophrenics, mental defectives and normal adults. *Amer. J. Psychiat.*, 1956, 112, 803-807.
21. WILLIAMS, M. Psychophysiological responsiveness to psychological stress in early chronic schizophrenic reactions. *Psychosom. Med.*, 1953, 16, 456-462.

Received July 6, 1956.

CONCLUSION-DRAWING, COMMUNICATOR CREDIBILITY, AND ANXIETY AS FACTORS IN OPINION CHANGE^{1, 2}

BERNARD J. FINE

Boston University

H OVLAND AND MANDELL (8), investigating the relative effectiveness of communications whose facts and arguments are the same but whose issue-related conclusions are or are not explicitly stated, found an explicit statement of the conclusion more effective than no statement of the conclusion in stimulating opinion change. Thistlethwaite, de Haan and Kamenetzky (19) suggest that greater opinion change might be associated with the explicit communication primarily because it produces greater comprehension of the communicator's position. Although their explicit communication produced greater comprehension of the communicator's position, they found no difference in opinion change between their explicit and nonexplicit groups. However, a direct test of their implicit assumption of a positive relationship between knowledge of the conclusion and opinion change was not made. In view of the inconsistencies between the findings of these two studies, a further examination of the potential effectiveness of conclusion-drawing seemed warranted.

Hovland and Mandell also explored the interaction of source credibility with conclusion-drawing. Their predictions that a noncredible communicator would be more effective using a nonexplicit communication and that a credible communicator would be more effective with an explicit communication were not substantiated. However, they indicate that their induction of source credibility may not have been sufficiently potent. Accordingly,

¹ The research is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Boston University. The author is now at the Army Quartermaster Research and Development Center, Natick, Massachusetts. He wishes to thank Dr. Walter Weiss, his major professor, for his invaluable counsel in all phases of the study and Dr. Nathan Maccoby, his second reader, for his many contributions, especially those which made this study possible in the first place.

² The research was carried out under contract ONR 492(04) between the Office of Naval Research and Boston University. The program of research is under the general direction of Dr. Weiss.

a second purpose of the current research was to test these predictions using a different induction procedure.

Since initial position (7, 11, 19) and audience anxiety (9, 10) have been shown to be relevant to opinion change, the interactive effects of both of these predispositional factors with conclusion-drawing were also examined. It was predicted that Ss initially opposed to the communicator's position would be more influenced by a nonexplicit communication whereas Ss initially favorable would be more influenced by an explicit one. With regard to anxiety, it was predicted that Ss high in anxiety would be more influenced by the anxiety allaying communication than would Ss low in anxiety. Furthermore, while it was felt that the explicit communication would be more effective than the nonexplicit one at both anxiety levels, it was predicted that the advantage of the explicit over the nonexplicit communication for Ss high in anxiety would be greater than for those low in anxiety.³

METHOD

Before Measures

The before questionnaire was presented to the Ss ostensibly as an opinion survey conducted by a national research organization. It contained four opinion items on biological warfare and a number of "filler" items on unrelated opinion issues. All items were multiple-choice in form. Also included were 20 items from the Taylor Manifest Anxiety Scale,⁴ presented as a "personal inventory."

Experimental Communications and After Measures

Three weeks after completing the before questionnaire, the Ss read an article on biological warfare presented as part of a Boston University study investigating how well people learn from newspaper articles. The communication, following the "logical" form employed by Hovland and Mandell (8), took the position that biological warfare is not a "super" weapon and that we in the United States can defend against it. In

³ For a more detailed statement of the assumptions underlying these predictions, see (3).

⁴ The twenty items were indicated by Harold P. Bechtoldt as being the most discriminating of all Taylor Scale items. Personal communication, July, 1955.

the explicit version, the final paragraph stated the conclusion that biological warfare is not a "super" weapon for us since we in the United States can defend against it. In the nonexplicit version, in place of the explicit statement of the conclusion, the Ss were urged to draw their own conclusions on the basis of the arguments presented in the article. The two communications differed in no way except for the last paragraph.

One half of the Ss read the explicit communication and one half read the nonexplicit. One half of the Ss in each of these groups had the authorship attributed to the New York Times, and the other half to the Daily Worker. The name of the source was mentioned both at the beginning and at the end of each article. For the Daily Worker, the explanatory phrase "official newspaper of the Communist Party in the United States" was stated at the beginning of the article. For the New York Times, the descriptive phrase "one of the most widely read newspapers in the U. S." was added. The four forms of the communication (conclusion-drawing $\times$ credibility) were distributed in accordance with a sequence of 4×4 Latin Squares so that no two adjacent Ss were assigned to the same experimental condition. The immediate-after questionnaire, in addition to containing the same four opinion items on biological warfare as the before questionnaire, also required the Ss to state the main position of the communicator and to make judgments concerning the article itself. Additional items measured acquisition of the content of the article, recall of source, and concern about biological warfare. Six weeks later the Ss were presented with the identical questionnaire ostensibly to "find out how well people remembered what they had read." After these questionnaires were collected, the Ss were informed of the true nature of the experiment.

Control Group

This group consisted of 36 students who were not exposed to the communications, but who completed appropriate questionnaires at the same times as the experimental groups. The control group was a single class used as a unit.

Subjects

Approximately 400 students, predominantly freshmen, at American International College, Springfield, Massachusetts, were used. The research was carried out in regularly scheduled class meetings in various psychology courses.⁵

Scoring

Each item on biological warfare was scored on a three-point scale. A response to an item indicating that biological warfare was not a "super" weapon was given a score of one; the "uncertain" category, a score of two; and response indicating that biological warfare was a "super" weapon, a score of three. The scores on the four opinion items were summated for each S for

each administration, yielding a theoretical range from 4 to 12. To exhibit change of opinion from any one administration to another, changes in total scores between administrations were computed for each S, a plus change score indicating that the S had changed toward greater acceptance of the position that biological was a "super" weapon. The theoretical range of change scores was from minus eight to plus eight.

The anxiety items were scored by assigning a score of one to a response indicating that the S was anxious and a score of zero to a response indicating that he was not anxious. The Ss scoring in the upper and lower quartiles of the total distribution of anxiety scores were designated as high anxious and those in the middle two quartiles as low anxious.

In previous studies involving the Taylor scale, high scorers have typically been compared with low scorers, the comparison being based on the assumption that the extent of manifest anxiety as measured by the scale is directly and linearly correlated with the level of the internal anxiety drive (16, 17, 18). However, Eriksen (1, 2) has suggested that the Taylor scale may rather be an indicator of the ways in which individuals handle their anxiety, a view congruent with the lack of differential conditioning or learning effects found by some researchers between Ss scoring high and Ss scoring low on the Taylor scale (4, 6, 13, 14, 15). If a distinction is made between level of internal anxiety drive and the overt responses made to the cues from that drive, it thus seems reasonable to assume, following Eriksen, that high and low scores on the Taylor scale reflect differences in modes of response to anxiety rather than in level of drive, and that those scoring in the middle of the distribution derived from the Taylor scale are less anxious than those with high or low scores. This assumption is supported by the finding of a curvilinear relationship between learning and Taylor scores (14, 15), with Ss scoring high or low performing similarly to each other but differently from those scoring toward the middle of the distribution.

RESULTS

Table 1 shows the mean opinion change scores from before to immediate-after for the four experimental groups and the control group. All of the experimental groups showed a significantly greater amount of opinion change than did the control group ($p < .01$, two tails for each comparison). The experimental groups had not differed from the control group initially.

The total group given the explicit treatment showed a tendency toward greater opinion change than the nonexplicit group ($p = .10$, two tails by analysis of variance⁶). There was no significant difference between the credible

⁵ The author wishes to thank Dr. Dorothy T. Spoerl and Dr. Charles A. Wells for the use of their classes. In addition, he is grateful to them and to Mr. Norman H. Berkowitz for assistance in administering certain phases of the research.

⁶ Initial differences in opinion between the explicit and nonexplicit groups suggested an analysis of covariance. Such an analysis could not be undertaken however, as a result of heterogeneity of regression between the groups.

TABLE 1

MEAN OPINION CHANGE SCORES FROM BEFORE TO IMMEDIATE-AFTER

Control Group		Experimental Groups			
		EC*	Ec	eC	ec
Means**	+.44	-2.26	-2.01	-1.68	-1.79
N	36	84	87	85	83

* E indicates explicit statement of conclusion, e nonexplicit; C indicates credible source, c noncredible.

** + indicates movement away from the communicator's position; - indicates movement toward the communicator's position.

and noncredible conditions in amount of opinion change from before to immediate-after.

The first item in the after questionnaire required the Ss to state the communicator's position. In the explicit condition 62 per cent of the Ss reported the communicator's position correctly, compared with only 40.5 per cent of the Ss in the nonexplicit condition ($p < .001$, two tails). Incorrect responses to this question were mainly restatements of some of the general principles and specific considerations that the article suggested were necessary to arrive at a rational decision concerning the dangers of biological warfare.

To test the possibility that the difference in opinion change between the explicit and nonexplicit groups is due to the difference between them in *expressed* knowledge of the communicator's position, the relationship between *expressed* knowledge of the communicator's position and opinion change was examined. No significant difference in opinion change was found between those expressing knowledge of the communicator's position and those not expressing such knowledge.

Initial Position

The Ss were divided at the median (6.47) into "favorable" and "unfavorable" groups in terms of their initial opinion scores. The mean opinion score for the favorables was 5.03; the mean for the unfavorables was 8.91. The unfavorable group showed greater opinion change than the favorable group ($p < .01$, two tails). The smaller opinion change of the favorable group is obviously related to the limited extent of change possible for those closer to the communicator's position. Within the favorable group the explicit treatment was more effective than the nonexplicit ($p = .02$, one tail).

TABLE 2

MEAN OPINION CHANGE SCORES FROM BEFORE TO IMMEDIATE-AFTER FOR Ss INITIALLY HIGH OR LOW IN ANXIETY*

	EC	Ec	eC	ec	All Groups
Means for low anxious Ss	-2.36	-1.74	-1.40	-1.38	-1.73
N	50	38	42	53	183
Means for high anxious Ss	-2.12	-2.22	-1.95	-2.37	-2.15
N	34	49	43	30	156

* High anxiety Ss are those scoring in either the upper or lower quartiles of the distribution of Taylor scale anxiety scores; low anxiety Ss are those scoring in the middle two quartiles.

Within the unfavorable group there were no differences between the explicit and nonexplicit treatments.

Anxiety

The obtained range of scores on the Taylor scale was from 0 to 18 and the obtained mean was 6.04. While the distribution shows some skewness, this is not inconsistent with the distributions obtained by others using the scale on college populations (12, 22). The abridged Taylor scale had an odd-even reliability of .78 using the Rulon formula (5). Table 2 gives the mean opinion change scores from before to immediate-after for high and low anxious Ss by experimental conditions. The total high anxious group shows a greater mean opinion change than does the total low anxious group ($p = .05$, one tail).⁷

The communication was equally effective for all conditions within the high anxious group; no opinion change differences were found between the explicit and nonexplicit and the credible and noncredible conditions. For low anxious Ss, however, the explicit treatment was more effective than the nonexplicit ($p = .03$, one tail).

Acquisition of Content

To check the comparability of the groups with respect to acquisition of content of the communication, fact-quiz scores for the experimental groups were compared. (The fact-

⁷ The separate mean change scores for the upper and lower quartiles of the anxiety distribution were greater than that for the combined middle two quartiles. However, only the Ss in the upper quartile changed significantly more than those scoring in the middle two quartiles.

quiz consisted of six multiple-choice items; the mean for all Ss was 5.09). The Ss reading the explicit communication show a significantly higher mean fact-quiz score than do the Ss reading the nonexplicit communication ($p = .06$, two tails). However, no relationship between acquisition of content (fact-quiz) and opinion change was found.

Judgmental Items

There were no significant differences between the credible and noncredible or the explicit and nonexplicit groups in the proportion of Ss saying that the article was fair or a piece of propaganda. Responses to both of the relevant judgmental items indicated that approximately 70 per cent of all the Ss felt that the article presented a fair treatment of the topic and was not propagandistic in intent.

Recall of Source

An open-ended question asking the S to write the name of the source was included in the after questionnaire. Table 3 shows the percentage of Ss in each experimental group recalling the source correctly. The Ss in the credible condition were more likely to recall the source correctly than were the Ss in the noncredible condition ($p < .02$, two tails). There was no significant difference between the explicit and nonexplicit groups in the proportion of Ss recalling the source correctly. No relationship was found between opinion change and recall of source.

Concern About Biological Warfare

After reading the communication, the Ss were asked two questions regarding their concern about biological warfare ("How concerned are you about biological warfare?") and the importance of the topic to them ("How important is the topic of biological warfare to you?"). Because the responses to the two questions were very highly correlated, the Ss were divided into two groups; those that had re-

TABLE 3
PERCENTAGE OF Ss RECALLING SOURCE CORRECTLY

	Experimental Groups			
	EC	Ec	eC	ec
Percentage	78.5	64.4	76.5	69.9
N	84	87	85	83

TABLE 4
MEAN OPINION CHANGE SCORES FROM BEFORE TO IMMEDIATE-AFTER FOR Ss RELATIVELY HIGH AND LOW IN CONCERN ABOUT BIOLOGICAL WARFARE

Degree of Concern	Experimental Group				
	EC	Ec	eC	ec	All Groups
Highly Concerned (N)	-.64 (11)	-1.46 (13)	-.82 (22)	-1.67 (24)	-1.20 (70)
Not highly Concerned (N)	-2.51 (73)	-2.11 (74)	-1.98 (63)	-1.76 (59)	-2.11 (269)

sponded that they were "extremely" or "very" concerned about biological warfare and who also felt that the topic was "extremely" or "very" important to them were designated as highly concerned and all other Ss as relatively nonconcerned. Opinion changes between and within the two groups are shown in Table 4. The Ss who indicated relatively low concern about biological warfare, after reading the communication, showed a greater mean opinion change than those who expressed high concern ($p < .01$, two tails). The two groups did not differ in initial opinions.

The Ss in the nonexplicit treatment were more likely than the Ss in the explicit treatment to say that biological warfare was of concern and importance to them ($p < .01$, two tails).

Delayed-After Results

After six weeks, the mean opinion score of each experimental subgroup mentioned in the analysis of the immediate-after data was still significantly different from the mean of the control group ($p < .01$, two tails). With respect to the comparisons between experimental subgroups, only the difference between the high and low concern groups approaches significance ($p = .11$, two tails). The latter group showed greater retained opinion change than the former.

At this time, the Ss in the credible condition were more likely than those in the noncredible condition to say that the author gave a fair presentation ($p < .05$, two tails); and there was a greater tendency for the Ss who read the nonexplicit communication than for those who read the explicit one to indicate that the author's treatment of the topic was fair ($p < .10$, two tails).

DISCUSSION

The indication that the explicit treatment is somewhat more effective than the nonexplicit tends to support the findings of Hovland and Mandell (8). However, the basis for the potentially greater effectiveness is not certain. Under certain conditions, knowledge of the communicator's position may be an important factor as Thistlethwaite, de Haan and Kamenezky (19) suggest. In the present research, however, no relationship was found between *expressed* knowledge of the communicator's position and opinion change. This finding is consistent with a conceptualization of opinion change as a continuous process in which the *S*, while reading the communication, is being affected by many content cues. While some people may not be able to order verbally everything they have read in a communication in a fully organized, meaningful, and closed form, they are apparently able to get enough of the "drift" of the communication to be as influenced by it as are others who can explicitly express the communicator's position.

To account for the superiority of the explicit condition over the nonexplicit in changing opinions, different assumptions are needed. Given a fairly complex communication, it is quite likely that the person, while reading, does not adequately perceive the logical organization of the content or arrange the data and arguments to form the implied syllogistic pattern represented by the current communication. The more complex the communication, the more likely he will have difficulty in grasping this informative patterning. Reading the conclusion may instigate awareness of the relationship existing between the preceding factual and logical groundwork and the basic point of the message. Thus, the explicit presentation of the conclusion may serve essentially to unify the body of the communication so that it becomes more clearly related to the conclusion. The impact of this unified, better understood communication is likely to be greater than in the nonexplicit condition where the person, if he has inferred the desired conclusion, may well be less certain of its appropriateness and validity. While it is probable that the person who arrives at the appropriate conclusion on his own achieves a degree of unification in the content of the message, the attendant uncertainty should serve to reduce

the over-all impact of the communication. Unfortunately, no measure of certainty was included in this study.

If the recipient of a communication about a potentially anxiety arousing topic is uncertain concerning its anxiety allaying conclusion, however, somewhat greater residual concern would be expected than if he were more certain about it. In line with this expectation, the data show that *Ss* in the nonexplicit condition evidenced more concern than the *Ss* in the explicit condition. If uncertainty of the correct conclusion underlies residual concern, then less opinion change may be expected among the highly concerned *Ss* than among those not as concerned. Again, the data bear this out (see Table 4).

The lack of opinion change differences between the credible and noncredible conditions may have been due to a weakness in the induction of the variable, but this possibility seems unlikely. It had not been considered necessary to obtain measures of the credibility attributed to the two sources by the *Ss*, since previous experience with other *Ss* indicated that the New York Times and the Daily Worker were perceived as differing greatly in credibility. To heighten this difference, qualifying phrases amplifying the credibility or lack of credibility were attached to the names of sources, it will be remembered, in the present study.

The results for initial position were possibly complicated by the generally favorable initial opinions of *all* the *Ss*. This fact may have additional consequences if it is assumed, as seems reasonable, that persons form an accurate impression of a communicator's position more readily when it is discrepant from their own. Here, even in the nonexplicit condition, *Ss* may be able to sense the main tenor of the communication because its message is so strongly opposed to their own position. But when the position of the *S* is not very discrepant from the communicator's, a finer discrimination is required. Thus, it might be expected that an explicit statement of the appropriate conclusion would result in greater opinion change for the initially favorable *Ss* than would no explicit statement.

The finding that *Ss* high in anxiety were more influenced by the communication than were those low in anxiety is consistent with the results of previous research (20, 21) in which

congruency between relevant response predispositions and the opinion response urged by the communicator were found to facilitate communication effectiveness.

SUMMARY

The study was designed to explore the simple and interactive effects of conclusion-drawing, communicator credibility, initial position, and anxiety. A before-after design was employed. The Ss' initial positions on biological warfare were assessed and they were given the Taylor Manifest Anxiety Scale. Three weeks later all Ss read an article on biological warfare which took the position that biological warfare was not a "super" weapon. This was followed immediately by another opinion questionnaire containing opinion items, items testing retention of content of the communication and items pertaining to the Ss' judgments of various aspects of the communication. Six weeks later all Ss took identical questionnaires. A control group that was not exposed to the experimental communication but took all of the questionnaires was employed. Some of the major findings are: (a) No relationship was found between *expressed* knowledge of the communicator's position and opinion change; (b) the explicit treatment was somewhat more effective in changing opinions than was the nonexplicit; (c) no difference was found in opinion change between the groups given credible and noncredible treatments; (d) the opinion change of Ss high in inferred anxiety was greater than those low in anxiety; and (e) Ss high in residual concern changed less in opinion than Ss not high in residual concern.

REFERENCES

1. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561-566.
2. ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety Scale and the Hysteria-Psychasthenia Scale from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135-138.
3. FINE, B. J. Conclusion-drawing, communicator credibility and anxiety as factors in opinion change. Unpublished doctor's dissertation, Boston Univer., 1956.
4. FRANKS, C. M. Conditioning and personality. *J. abnorm. soc. Psychol.*, 1956, **52**, 143-151.
5. GUILFORD, J. P. *Psychometric methods* (2nd ed.). New York: McGraw-Hill, 1954.
6. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94-99.
7. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. New Jersey: Princeton Univer. Press, 1949.
8. HOVLAND, C. I., & MANDELL, W. An experimental comparison of conclusion-drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.*, 1952, **47**, 581-589.
9. JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504-518.
10. JANIS, I. L. Anxiety indices related to susceptibility to persuasion. *J. abnorm. soc. Psychol.*, 1955, **51**, 663-668.
11. JANIS, I. L., LUMSDAINE, A. A., & GLADSTONE, A. I. Effects of preparatory communications on reactions to a subsequent news event. *Publ. Opin. quart.*, 1951, **15**, 487-518.
12. KERRICK, JEAN S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1955, **50**, 75-78.
13. KERRICK, JEAN S. The effects of manifest anxiety and IQ on discrimination. *J. abnorm. soc. Psychol.*, 1956, **52**, 136-138.
14. MATARAZZO, RUTH G., & MATARAZZO, J. D. Anxiety level and pursuitmeter performance. *J. consult. Psychol.*, 1956, **20**, 70.
15. MATARAZZO, J. D., ULETT, G. A., GUZE, S. B., & SASLOW, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, **18**, 201-206.
16. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.
17. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-89.
18. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
19. THISTLETHWAITE, D. L., DE HAAN, H., & KAMENETZKY, J. The effects of "directive" and "nondirective" communication procedures on attitudes. *J. abnorm. soc. Psychol.*, 1955, **51**, 114-119.
20. WEISS, W., & FINE, B. J. Opinion change as a function of some intra-personal attributes of the communicatees. *J. abnorm. soc. Psychol.*, 1955, **51**, 246-254.
21. WEISS, W., & FINE, B. J. The effect of induced aggressiveness on opinion change. *J. abnorm. soc. Psychol.*, 1956, **52**, 109-115.
22. WOLFF, W. M. Certainty: generality and relation to manifest anxiety. *J. abnorm. soc. Psychol.*, 1955, **50**, 59-65.

Received July 3, 1956.

THE ROLE OF EXPERIENCE IN JUDGING CHILDREN'S PHOTOGRAPHS

WILLIAM KESSEN¹

Yale University

IN a recent review of the psychological literature on the perception of people, Bruner and Tagiuri conclude: "The first step in reacting to another is forming an impression of him. Later reactions depend on this first step. If there is to be a science of interpersonal behavior, it will rest upon a cornerstone of social perception" (1, p. 650). This assertion is of special practical and theoretical importance when the perception of children is under consideration. For one, whatever impressions a parent forms of his child will affect, in as yet undetermined ways, subsequent interactions between parent and child, and it is important to determine to what degree variations in the perception of children can be ascribed to the stimulus presented or, failing that, to determinants in the personality of the parent. An approach to this problem can be made by evaluating the *consistency* with which different parents judge the same child stimulus. Moreover, by comparing with this evaluation some independent index of *validity* there is the additional gain of determining whether parents, when compared with nonparents, have a greater or lesser accuracy of judgment as a result of their intimate contact with at least one member of the class of child stimuli.

Problems of much the same order can be stated for the professional student of human behavior, the psychologist. Does his training or clinical experience modify either the consistency or the validity of his perceptions of people? The judgment of children, particularly in the younger range, presents a case of special importance, because many of the cues for personality evaluation available from adult human beings are not present, perhaps forcing the psychologist back to what some writers have called "physiognomic perception" (5, 6).

The present study was designed as an initial

approach to this area of problems by comparing the consistency and validity of judgments made of children's photographs, using judges who varied in their experience with, or knowledge of, children.

METHOD

From the photographic files of the Yale Child Study Center longitudinal study (2, 7), glossy photographs of 10 children (6 girls and 4 boys) were selected. At the time the pictures were taken, the children ranged in age from 14 to 23 months, the median age being 19 months. The pictures were full length and the children were naked. When several photographs of a single child were available, selection was made on the criteria of (a) being as near full-face to the camera as possible, and (b) reduction of extraneous objects (toys, mothers' hands, etc.). The 10 photographs selected were reprinted to a common size (4" x 8"), mounted on heavy white cardboard, 8" x 10", and ordered alphabetically by the code names of the children.

This series of photographs was presented to 50 Ss in individual sessions. The Ss fell into three groups:

1. Undergraduates: Twenty male undergraduates, all childless.

2. Parents: Ten junior faculty members and their wives, who met the additional requirements that none was a psychologist, and that they had a child at least as old as those shown in the photographs, but no more than one year older.

3. Experts: Ten professional psychologists, either faculty members or senior graduate students. Three of these also met the "relevant child" requirement imposed on the parent group.

The instructions given to all Ss were as follows:

I'm interested in how well people are able to tell something about children on the basis of a single photograph. There are ten pictures in the series, of six girls and four boys, just past their first birthday. What I want you to do is rank them, that is, put them in order, for each of several traits.

For example, how would you rank these pictures as showing *intelligence*—how bright the child is.

Now, would you put them in order again, this time as a function of *sociability*—how well they react to and interact with people.

The next characteristic I'd like you to sort on is *activity*—how vigorous and active the child is.

And now, could you rank the pictures on *sensitivity*—that is, how readily the child responds to *things* going on around him or her.

And finally, I'd like you to judge *mental health*—the order of good functioning or healthy adjustment.²

¹ The writer is indebted to Professor Katherine M. Wolf who suggested the research and made the "contemporary validity" judgments, and to Dr. Sally Provence, who made the "predictive validity" judgments.

² All Ss were also asked to select the most attractive and least attractive boy and girl. Because an ordering

At the end of each sort, *E* recorded the *Ss*' ranks and returned the pictures to their original order before presenting them again. No special restrictions were put on methods of handling the mounted photographs, but almost all *Ss* spread the pictures out and then indicated their rank order by serializing the pictures spatially or restacking them in the chosen order. If any *S* made comments or asked questions about the instructions, *E* made some nondirective comment, or accepted without comment the definition the *S* set for the task. In some cases when the *S* persisted in questioning the task (almost exclusively for the Sensitivity ranking), *E* gave a paraphrase of the original instructions.

As an index of validity, a child psychologist who had observed the children intensively through their first two years, made rankings on the same instructions, with a further request to make her judgments relevant to the first half of the children's second year. She was not shown the photographs but rather small white cards on which the code names of the children were typed.

To obtain a "prediction" index, a pediatrician of the longitudinal study also ranked the cards on the instructions given earlier, but with the qualification that her judgments be on the basis of the last time she had seen the child. Because of attrition in the study sample, this last contact ranged from age 16 months to age 5 years, with the median at age three and one-half years.

Beyond these judgmental indexes of validity, Hetzer-Wolf (3) test results were available for all children as a validity check of the Intelligence ranking.

RESULTS

Acceptance of the Task.

A majority of *Ss* in all groups met the initial instructions with the remark: "Impossible" or a similar reaction of doubt, and then proceeded with scarcely a pause to perform the task. No questions were asked about the Intelligence (*I*), Activity (*A*) or Health (*H*) instructions, and only two about the Sociability (*Soc*) ranking (both from Experts working on socialization). The Sensitivity (*Sen*) ranking caused a fairly large number of hesitations and questions, however, turning in a more or less sophisticated way on the issue of whether the instructions asked for *quality* of sensitivity (does the child respond adequately?) or for *thresholds* (is the child emotional, irritable?). This confusion is reflected in the results reported below.

Concordance among *Ss*.

The first questions put to the data were: How consistent are the *Ss*' responses, and are procedure different from that used with rankings was necessary to evaluate this rating, no results drawn from it are included in this report.

there variations in consistency assignable to groups? The index chosen was Kendall's coefficient of concordance *W* (4), and the values obtained are shown in Table 1. This coefficient indicates the degree to which *Ss* make the same rankings. Twelve of the 15 independent statistics shown in the upper lines of Table 1 are significant, with those for *A* and *Sen* being consistently lower than the other three rankings for all groups. Moreover, there seems to be no systematic variation among the groups, the Undergraduates being as consistent as either of the experienced groups. The variation in consistency shown, therefore, is assignable to differences among scales rather than among groups of *Ss*.

The degree of consistency shown in the upper lines of Table 1 was based on groups of 20, 20 and 10 judges. If, however, each group is seen as representing a single judge (by calculating mean ranks for each child for each group) and *W* recomputed on this basis, the values increase markedly in size (Table 1). The same comparisons are represented in the more usual form of correlation coefficients (Kendall's *tau*) in Table 2, showing the cor-

TABLE 1
CONCORDANCE (*W*) WITHIN GROUPS OF JUDGES
(In the Combined Case, Each Group is Treated as a Single Judge)

Group	Variable Ranked					
	<i>N</i>	<i>I</i>	<i>Soc</i>	<i>A</i>	<i>Sen</i>	<i>H</i>
Undergraduates	20	.31***	.44***	.15**	.08	.44***
Parents	20	.23***	.48***	.18***	.06	.35***
Fathers	10	.26**	.48***	.12	.23*	.35***
Mothers	10	.26**	.52***	.33***	.08	.46***
Experts	10	.38***	.39***	.07	.27**	.39***
Combined	50	.92**	.90**	.85**	.65*	.88**

* $p < .05$.

** $p < .01$.

*** $p < .001$.

TABLE 2
CORRELATIONS (*tau*) AMONG GROUPS
(Each Group is Represented by Mean Values)

Group	Variable Ranked				
	<i>I</i>	<i>Soc</i>	<i>A</i>	<i>Sen</i>	<i>H</i>
Parents-Underg.	.73**	.67**	.66**	.54*	.78***
Parents-Experts	.63**	.72**	.72**	.39	.64**
Underg.-Experts	.85***	.78***	.55*	.18	.60**

* $p < .05$.

** $p < .01$.

*** $p < .001$.

relation of the mean ranks of each group with those of the other groups.

These results indicate that, at least for the present task, a high level of consistency is obtained when each group is considered a single judge and compared with other groups, and a lower but significant level of consistency is obtained when individual judges are compared within groups. Moreover, the three groups do not vary in their consistency of judgment and are making, to a significant degree, the *same* rankings.

The Parent group can be dichotomized by sex to evaluate consistency variations on this dimension. Table 1 presents the obtained *W*'s, and except for the paradoxical finding that fathers are more consistent in the rankings of Sensitivity and mothers more consistent in the ranking of *A*, there are no large overall differences between male and female parents. Incidentally, it was found that husband-wife pairs do not show higher correlations of judgment than do unrelated men and women.

Halo Effects.

For each *S*, intercorrelations of each ranking with every other ranking were made, to determine the tendency of *S*s to make the same ranking of the pictures regardless of instruction, i.e., to assess halo effect. The 10 obtained values for each *S* were averaged without regard to sign. Table 3 summarizes these calculations, giving range and median *tau* value for each group on each ranking. As can be seen, the range of variation among groups is narrow, with only a slight difference separating the Undergraduates from the experienced groups. In effect, there is no consistent tendency for halo effect to vary among groups of judges.

Validity of Judgments.

Each ranking of the 50 judges on the five scales was correlated with the rankings made by the child psychologist who had studied the children intensively ("contemporary validity"). The results of these computations are shown in Table 4 which lists the median value within each group of the correlation coefficients (again, Kendall's *tau*). They are in general low, and no median value is significantly different from zero. If, however, a validity run analogous to the consistency run

TABLE 3
MEDIAN INTERCORRELATION (*tau*), AND RANGE OF INTERCORRELATION FOR EACH GROUP, OF ALL SCALES WITHIN SUBJECTS
(i.e., the Degree of Halo Effect)

Group	Median <i>tau</i>	Range for Group
Undergraduates	.41	.16-.65
Parents	.34	.17-.63
Experts	.32	.20-.72

TABLE 4
MEDIAN VALIDITY CORRELATIONS
(Each *S*'s Rankings are Correlated with the Criterion)

Group	Variable Ranked				
	<i>I</i>	<i>Soc</i>	<i>A</i>	<i>Sen</i>	<i>H</i>
Undergraduates	.39	.29	.09	-.11	.24
Parents	.22	.38	.29	-.11	.24
Experts	.24	.38	.13	-.07	.29

of Table 2 is made, that is, considering each group to be a single judge, much higher coefficients are obtained (Table 5). Not only are the rankings consistent when combined, but they also become clearly more valid. Of the 15 independent statistics, 9 are significantly different from zero and in the proper direction, and two (Undergraduates-*Soc* and Experts-*H*) are barely short of the five per cent level. The *Sen* index is once more deviant, even showing one significant correlation of negative sign. However the comparisons are made, there are no gross differences among groups. If the *Sen* correlations are withdrawn as belonging somehow to a different population, there is a slight tendency for the mean correlations (*tau*) from Table 5 to favor the Parents (.52) over the Undergraduates (.46) and the Experts (.54) over the Parents.

The correlations with the validity criterion of the male parents considered as a single judge and the female parents so considered are also shown in Table 5, a further demonstration that differences on this dimension are small and apparently unsystematic.

To appraise "predictive validity" the rankings of the groups considered as three single judges were correlated with the rankings of the staff pediatrician who had examined the children throughout their stay in the longitudinal research, and who made her judgments relevant to the last time the children were seen. Table 6 shows that the relation-

TABLE 5

CORRELATIONS (*TAU*) WITH VALIDITY CRITERION
(Each Group Considered as a Single Judge)

Group	Variable Ranked				
	<i>I</i>	<i>Soc</i>	<i>A</i>	<i>Sen</i>	<i>H</i>
Undergraduates	.56*	.42	.39	-.33	.47*
Parents	.56*	.49*	.56*	-.54*	.47*
Fathers	.38	.57*	.42	-.36	.47*
Mothers	.54*	.49*	.60**	-.20	.38
Experts	.54*	.69**	.49*	-.27	.42

* $p < .05$.

** $p < .01$.

TABLE 6

CORRELATIONS (*TAU*) WITH PREDICTION CRITERION OF
VALIDITY
(Each Group Considered as a Single Judge)

Group	Variable Ranked				
	<i>I</i>	<i>Soc</i>	<i>A</i>	<i>Sen</i>	<i>H</i>
Undergraduates	.24	.47*	-.11	-.14	.47*
Parents	-.07	.22	-.02	-.40	.33
Experts	.18	.51*	.04	-.45	.29

* $p < .05$.

ships decay to a point just short of chance expectancy, even though the correlation (*tau*) between the two validity judges is marked and consistent (*I*—.42, *Soc*—.47, *Sen*—.60, *H*—.73, all values but the first significant at the .05 level or better). In a word, the validity of judgment by the *Ss* is higher for "contemporary" judgments than for "predictive" ones, although *Soc* and *H* correlations in the latter case are sometimes significant.

A final index of validity is provided by the correlations of the *I* rankings of the groups, considered as three single judges, with scores on Hetzer-Wolf (3) tests given within six months of the time each child was photographed. For the Undergraduates and the Parents, the correlation was .47 and for the Experts, it was .54, values significant at the .05 level. The comparable values for contemporary validity were .56, .56 and .54 (Table 5). This finding is particularly important in demonstrating that the validity judge did not merely guess accurately at the expected rankings of the *Ss*.

DISCUSSION

Among the findings obtained in this study, those which warrant some special mention are

(a) the high consistency and validity of groups as judges compared with individuals as judges, and (b) the failure to show variations in consistency or validity on the dimension of experience with children, practical or professional.

The first problem seems to be answered in a rather straightforward way. By pooling judgments there is a tendency for idiosyncratic variations to be eliminated, much as the probability of a particular die face appearing is more stably measured with an increasing number of rolls. One source of such idiosyncratic responding (probably not random) in this study was expressed explicitly by a number of parents who remarked on the similarity of one or another of the photographs to their own children. Presumably other such biases, different in different subjects, operated to introduce error into the rankings of individual judges.

This line of discussion leads to a more difficult problem—the assumption that there is something stable in the situation for the errors to distort, something analogous to a particular die face. Nothing in the results presented here contradicts such an assumption, although the correlations of *Ss*' rankings with the validity rankings are modest, even when statistically significant.³ To get at this something which is perceived by the *Ss*, one can make two attacks, either a systematic variation in cues presented (changes in the photographs; for example, restriction to face alone), or an examination of response differences among individuals or groups. This second goal might be achieved by interviewing "good" and "bad" judges, or by comparing groups of judges who differ in some definable way. It was the latter procedure which was followed in the present study. The results of comparing groups selected to vary in their experience with children were disappointing insofar as they showed no systematic differences in any characteristic (concordance, halo, validity) which could be

³ This statement is mitigated somewhat by the consideration that the photographs used were not, like those in many social perception studies, posed; the children were in more ways than one *au naturel*. However, the fact that the validity correlations are substantially lower than the consistency correlations points to the danger of substituting implicitly the one for the other in social perception research. Agreement does not necessarily indicate accuracy.

assigned to group membership. It seems therefore that this first step in the analysis of the perception of children has produced a weak conclusion—that the characteristics measured do not vary with experience of the judges with the class of objects judged.

SUMMARY AND CONCLUSIONS

Still photographs of ten children in their second year were presented to 3 groups of Ss (20 undergraduate men, 20 parents and 10 professional psychologists) in individual sessions. All Ss were instructed to rank the photographs on Intelligence, Sociability, Activity, Sensitivity, and Mental Health. Validity of the rankings was indicated by correlation with the rankings of a child psychologist who had observed the children closely throughout their first two years.

The results obtained appear to warrant the following conclusions:

1. All groups of Ss made consistent judgments of Intelligence, Sociability, and Health. Concordance measures for Activity and Sensitivity were lower for all groups.
2. This consistency did not vary with experience with children, defined either on a practical (Parents) or on a professional training (Experts) basis.
3. For the present task, groups of Ss combined to provide a single measure were highly consistent with other such combined groups.
4. Median validity correlations of individual

Ss on all rankings were low, and not statistically significant.

5. When the groups were combined to function as single judges, validity on four of the rankings (all but Sensitivity) was high and, in nine of twelve correlations, statistically significant.

6. There were no striking differences in validity, however determined, among the three groups, or between male and female parents.

7. These findings suggest that whatever perceptual variables underlie judgments of children's photographs, these variables are not demonstrated to be related to practical experience or to professional training.

REFERENCES

1. BRUNER, J. S., & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison Wesley, 1954.
2. COLEMAN, R. W., KRIS, E. & PROVENCE, S. The study of variations of early parental attitudes. *Psychoanal. Stud. Child*, 1953, **8**, 20-47.
3. HETZER, H., & WOLF, KATHERINE. Babytests. *Z. f. Psychol.*, 1928, **107**, 62-104.
4. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
5. KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt Brace, 1935.
6. WERNER, H. *Comparative psychology of mental development*. Chicago: Follett, 1948.
7. WOLF, KATHERINE M. Observation of individual tendencies in the first year of life. In M. J. E. Senn (Ed.), *Problems of infancy and childhood*. New York: Macy, 1953.

Received July 9, 1956.

RELATIONSHIPS BETWEEN CHILD TRAINING AND THE PICTORIAL ARTS¹

HERBERT BARRY, III

Yale University

An artist's personality is commonly thought to be expressed in the style of his art creations, so that individual differences in art style are related to individual differences in personality. If this assumption is true, it should be possible to find a correlation between an appropriate index of personality and art style. The index of the artist's personality may include events known to be influential, such as the socialization practices he has undergone. A correlation between socialization and art style, if obtained, may mean that the socialization and art variables are related to the same aspect of the artist's personality.

This argument may be generalized from the single individual to the typical or modal individual in a small and relatively homogeneous society. The socialization practices and art styles as aspects of culture are perpetuated beyond the individual's life span by custom, but cultural custom like an individual's behavior develops and may be modified as an adjustment to the personality characteristics of the individuals. The present study was carried out to test for a correlation between severity of socialization and style of art among a sample of nonliterate societies. The measures on socialization have been presented by Whiting and Child (4), and the present investigator independently obtained measures of art style.

METHOD

Thirty nonliterate societies were selected from the list of 76 on which Whiting and Child gave socialization data. These 30 societies comprised all those from which the investigator was able to find at least ten works of graphic art, either exhibited in museums, or as illustrations in ethnographic reports. If 10-20 works of art were available for a culture they were all rated; if more

than 20 were available, representative ones were picked at random. In all, 549 works of art were rated.

Each work of art was rated on 18 criteria of art style, using a 7-point scale. Since the score on most of the variables was influenced by the material used, the art works of each culture were compared with those of other cultures in the same material, e.g., weaving, woodwork, wickerwork, and pottery. The culture was designated as above or below the median in each variable, combining its score in the different materials but comparing its art in each material only with the art of the other cultures in the same material.

Whiting and Child described five universal systems of behavior: oral, anal, sexual, dependence, and aggression. Two stages of socialization were distinguished for each system: initial satisfaction and socialization anxiety. The resultant 10 variables had each been rated on a 1-7 scale by three judges, and the three ratings added together. This combined score for each of the 10 variables was transformed into standard scores, using the sample of 30 cultures on which art scores were available. The 5 scores of initial satisfaction were reversed in sign so that a high score would mean low satisfaction. The 10 standard scores were then combined for an overall measure of severity of socialization, with a high score signifying high severity.

RESULTS

Eleven art variables were considered to be measures of complexity of art style.² One of these variables, entitled complexity of design, was defined at the upper extreme as a design with many unrepeatable figures to form a complex organization of design; the lower extreme of this variable was defined as a design with few figures or repetition of figures to form a simple organization of design. The other 10 variables, which were less directly defined in terms of complexity of art style, are listed in Table 1.

To provide a combined measure of complexity of art, cultures were scored as high in combined complexity if they were rated above the median (in the direction of complexity) in 6 or more of the 11 separate measures, low if they were above the median in 5 or fewer of these measures. Table 1 (left-hand column of figures) shows the tetrachoric correlation of

¹ This paper is a revised version of a thesis presented to the Social Relations Department at Harvard College in partial fulfillment of the requirements for the B.A. degree with honors. The thesis was written under the generous and helpful guidance of Dr. John W. M. Whiting. The revisions were made with helpful suggestions from Dr. Irvin L. Child.

² Seven additional variables were rated but are omitted here for the following reasons: 3 concerned color and there were insufficient cases; 2 were unreliably rated; 2 were rated with insufficient variation.

TABLE 1
CORRELATION OF THE ART VARIABLES WITH
COMPLEXITY OF STYLE AND SEVERITY OF
SOCIALIZATION

(All of the correlations are positive.)

Art Variables	Complexity of Style (r_{tet} , $N = 30$)	Severity of Socialization (r_{bis} , $N = 28$)
Complexity of design	.98**	.71**
Presence of enclosed figures	.91**	.32
Presence of lines oblique to each other	.91**	.13
Presence of sharp figures	.81**	.18
Presence of curved lines	.81**	.07
Representativeness of design	.67*	.56*
Presence of lines oblique to edges	.67*	.45
Crowdedness of space	.67*	.12
Asymmetry of design	.50	.26
Presence of border	.31	.20
Shortness of lines	.31	.11

* $p < .05$.

** $p < .01$.

each variable with this combined measure, using the table compiled by Davidoff and Goheen (1). The statistical significance was computed by the Fisher-Yates exact test (2).

The biserial correlation of each variable of art style with the overall measure of severity of socialization is shown in the right-hand column of Table 1. For each of the 11 art variables, the direction of high complexity is positively related to high severity of socialization. Two of the art variables are significantly related to the overall measure of socialization by t test. Complexity of design, which is most closely related to the combined measure of complexity of art, is the art variable showing the highest correlation with severity of socialization. The combined measure of complexity of art has a biserial correlation of $+ .47$ with the overall measure of severity of socialization ($t = 2.06$; $p < .05$).

The biserial correlation of complexity of design with each of the ten measures of socialization is shown in Table 2. In all 10 cases high severity of socialization is positively correlated with high complexity of design, although the intercorrelations among the 5 systems of behavior (reported by Whiting and Child) are mostly slight.

Table 3 lists the 30 cultures in this sample, grouped into those above and below the median on complexity of design. The left-hand column of figures for both groups of cultures

shows the combined measure of severity of socialization. The right-hand column indicates that complexity of design is correlated with the occurrence of cultures for which some of the socialization ratings were omitted because of insufficient information. Ratings were omitted in none of the 15 cultures above the median in complexity and in 7 of the 15 cultures below the median. This difference is significant by the Fisher-Yates Exact Test ($p < .01$).

Instances of severe socialization practices are generally more dramatic and likely to be reported more fully. Therefore it is probable

TABLE 2
CORRELATION OF COMPLEXITY OF DESIGN WITH
MEASURES OF SEVERITY OF SOCIALIZATION
(Satisfaction and anxiety measures taken from Whiting and Child (4))

Behavior system	Satisfaction		Anxiety	
	r_{bis}	N	r_{bis}	N
Oral	-.59**	29	+.35	27
Anal	-.32	23	+.16	23
Sexual	-.05	28	+.28	28
Dependence	-.46	27	+.42	26
Aggression	-.04	30	+.33	30

** $p < .01$.

TABLE 3
COMPLEXITY OF DESIGN AND SEVERITY OF
SOCIALIZATION IN THE SAMPLE CULTURES

Complexity of Design Above Median			Complexity of Design Below Median		
Culture	Socializ. severity	Not rated	Culture	Socializ. severity	Not rated
Ashanti	+1.01	0	Hopi	+.37	0
Chiricahua	+.77	0	Thonga	+.34	0
Dahomean	+.69	0	Navaho	+.20	0
Aloré	+.38	0	Paiute	-.05	0
Western Apache	+.35	0	Masai	-.28	4
Kwakiutl	+.29	0	Papago	-.30	0
Samoans	+.26	0	Ifugao	-.45	2
Ainu	+.12	0	Andamane	-.54	2
Maori	+.12	0	Marshall-ese	-.56	3
Marquesan	+.09	0	Chenchu	-.58	0
Arapesh	+.06	0	Yagua	-.62	0
Yakut	-.01	0	Comanche	-.94	0
Balinese	-.21	0	Murngin	-1.05	4
Trobrianders	-.22	0	Omaha	-	6
Teton Dakota	-.47	0	Zuni	-	8

Note:—Cultures are divided into above and below the median for the art variable of complexity of design. Each culture is listed with its average score on the combined measure of severity of socialization, and the number of socialization variables (out of 10) on which a rating was not made.

that omissions of ratings indicate an unreported low degree of severity more often than an unreported high degree of severity. There is evidence for this inference from the fact that the socialization ratings made with the lower of two specified degrees of confidence were generally below average in severity. If the majority of omissions indicate low severity of socialization, they give further evidence of the positive relationship described above with low complexity of art design. Since the art materials and socialization information were obtained from different sources, it is unlikely that the correlation indicates a direct causative relationship between inadequate ethnographic information and availability of art works below the median in complexity of design.

DISCUSSION

The correlation between severity of socialization and complexity of design in art works indicates the presence of a connecting link between these two variables, to which both are related. Cultures with low and high severity of socialization are certain to differ from each other in the typical personality characteristics of their people, if we believe that socialization has an important influence on personality and that our measures of socialization are valid. Therefore, cultures with simple and complex art style also differ from each other in the same typical personality characteristics of their people. It is reasonable to believe that these personality characteristics influence the style of the art works created by the people. This belief provides a possible explanation for the correlations between socialization and art variables.

The correlations of the 11 variables of art style with the overall measure of art style suggest the presence of a general factor of complexity or elaborateness in various specific measures of style. This general factor may be a component of the distinction between classicism and romanticism. In non-representative designs, which characterize the majority of the art works in this sample, romanticism might be described in terms of complexity of design and specific variables such as asymmetry, curved lines and crowded figures. Most of these 11 variables have been

included in a list of variables described by C. Sachs (3) as related to the classic-romantic distinction.

The oral and dependence systems of behavior, shown in Table 2 to have the highest relationships to complexity of design in art, are mainly concerned with adequacy of food, affection, and protection. In the majority of cultures with complex art style, to draw an implication from this finding, the typical individual learns self-reliant behavior to a high degree and is punished or frustrated for overt expression of dependence. On the other hand, the systems of behavior with lower relationships to art style (anal, sexual, and aggression) are concerned with prohibition or restriction of pleasurable actions, and imply pressure toward obedient and compliant behavior. The correlation of complex art style with severe socialization thus apparently applies primarily to severe socialization pressures toward independent behavior rather than toward obedient behavior.

The correlations reported in this paper are subject to various alternative interpretations. Such ambiguity can best be reduced by determining with more certainty the meanings of the measures of socialization and art style, as by finding the relationships of these measures to independent measures of personality, aesthetic expression, and other aspects of cultural practices.

SUMMARY

Works of pictorial art of 30 nonliterate cultures were rated on a number of variables of art style. Eleven of these art variables were found to be related to each other on a dimension of complexity of style. The art variables were correlated with a combined measure of severity of socialization from ratings presented by Whiting and Child (4). For each of the 11 art variables the cultures above the median in complexity tended to be above average in severity of socialization; for two of the art variables the correlation was statistically significant. The variable entitled complexity of design, which was more closely related than any other art variable to the overall measure of complexity, showed a higher correlation than any of the others with the measure of severity of socialization. Cul-

tures above the median in complexity of design tended to be above average in each of the 10 separate measures of socialization severity.

A possible interpretation of the results is that a personality characteristic is related to severity of socialization and complexity of art style in the individual, and that the cultural custom which perpetuates a pattern of socialization and art style is an adjustment to the personalities of the individuals in the culture.

REFERENCES

1. DAVIDOFF, M. D., & GOHEEN, H. W. A table for the rapid determination of the tetrachoric correlation coefficient. *Psychometrika*, 1953, **18**, 115-121.
2. FINNEY, D. J. The Fisher-Yates test of significance in 2×2 contingency tables. *Biometrika*, 1948, **35**, 145-156.
3. SACHS, C. *The commonwealth of art*. New York: W. W. Norton, 1946.
4. WHITING, J. W. M., & CHILD, I. L. *Child Training and personality*. New Haven: Yale Univer. Press, 1953.

Received July 10, 1956.

THE GENERALITY OF ACCURACY AND RESPONSE SETS IN INTERPERSONAL PERCEPTION¹

WAYMAN J. CROW AND KENNETH R. HAMMOND²

University of Colorado

PSYCHOLOGISTS have often expressed the belief that a successful adjustment to a social environment depends upon an accurate perception of other people in the environment. This perceptual process has been called "empathy," "social perceptiveness," "social sensitivity," and "insight into others," but will here be called *interpersonal perceptiveness*, according to current usage (cf. 6).

Previous students of this problem have followed formally similar procedures. Judges (subjects), after observation of some other individual or group (objects), are asked to estimate some characteristic of the objects³ for which some operational measure is available. This measure is then used as a criterion to determine the accuracy of the Ss. Although a wide variety of procedures have been used within this framework, three main types may be distinguished: (a) estimation of responses to personality inventories, (b) estimation of responses to questionnaires, and (c) estimation of ratings within small groups.

Studies using these techniques are indeed apparently similar—but do they produce similar results; i.e., are results generalizable? Taft (10) has pointed out the paucity of studies comparing these techniques, and Gage and Cronbach (6) note that many writers assume that a general ability or trait of interpersonal perceptiveness exists. If such a general trait does in fact exist, then results obtained using different techniques are comparable. If no such general trait is established, then greater caution than some writers have exercised will be necessary when results are compared.

¹ This investigation was carried out as part of a larger study of medical education supported by the Commonwealth Fund of New York.

² The authors are indebted to Fred Kern, Jr., M.D., for his assistance during many stages of the study.

³ To avoid confusion those persons who make judgments about others are called judges or subjects; those persons about whom judgments are made are called objects.

Moreover, any study of generality must take into account the fact that as knowledge has accumulated about the judgments that Ss make of others on the basis of brief observation, it has become apparent that what the S "brings to the situation" determines to a marked degree his response to the task. In attempts to sharpen concepts and to eliminate artifactual components of accuracy scores, attention has been called to the S's tendencies to "project" (8), to "assume similarity" (4), and to have "response sets" (3) and "implicit personality theories" (2). That such artifactual components invalidate conclusions regarding interpersonal perception accuracy has been pointed out in Campbell (1), Cronbach (2), and Crow (3). Such subject-determined (as distinct from situational- or object-determined) characteristics, in addition to their effect upon accuracy, may have psychological interest in their own right and may, in fact, be of more significance than accuracy in certain instances. For example, Fiedler has demonstrated the significance of "assumed similarity" for the prediction of team effectiveness (5).

This paper is a report of two studies. The purpose of the first study was to investigate the generality of interpersonal perceptiveness over *different procedures* when the influence of response sets upon the accuracy scores was controlled statistically. A variety of interpersonal perception measures was administered to a group of Senior medical students. Their scores on these different measures were intercorrelated in order to determine the extent to which generality over measures exists.

The purpose of the second study (involving different Ss) was to investigate the generality of the accuracy of interpersonal perception over *time* when the effects of response sets have been eliminated, and, in addition, to investigate the generality-over-time of certain response-set characteristics themselves. The same measures of interpersonal perceptiveness were administered three times, at six-month intervals, to a single group of Ss.

METHOD

The Scoring of Measures of Interpersonal Perception

Conventionally, interpersonal perception tests have been scored by finding the difference between the judge's estimate and the object's performance. These differences are then summed for all the estimations that the judge made about the object. Critical analysis of this procedure has revealed that such scores contain components that may have little to do with accuracy. Gage, Leavitt, and Stone (7) suggest the use of "intermediary scoring keys" and Cronbach (2) has proposed an analysis of variance method of scoring that eliminates the effects of these components upon the accuracy score and yields a highly refined measure of "differential accuracy." "Differential Accuracy reflects ability to predict differences between O's on any item" (2, p. 179). The author (3) has proposed a Random-Comparison Method of determining differential accuracy which differs from Cronbach's method in that it is distribution-free, or "nonparametric." Cronbach's method has the advantages of greater precision and scope whenever the assumptions of normality of distribution of the judge's estimations and the object's scores are warranted. The Random-Comparison Method has the advantage of usefulness when assumptions of normality are not justified. Although the latter method may be less time consuming to compute, this advantage is lost whenever it is desirable to obtain separate measures of the various response set components. This objective is accomplished easily with Cronbach's mathematical model, but requires separate computations with the Random-Comparison Method.

In the present study the Random-Comparison Method was used to eliminate artifactual components of the scores. Whenever the judges make estimations about several objects, a distribution of comparisons is generated by comparing every estimation with every object's score, ignoring the signs. Thus, if eight judges each make an estimation for each of ten objects, the comparison procedure would yield 800 difference scores. The difference score obtained by comparison of the judge's estimation with the score of the object for whom the estimation was intended can be compared with the median of the random comparison distribution. If this difference score is more accurate (lower) than the median of the random comparisons, then the judge has obtained differential accuracy with that object. The final score is the percentage of objects upon which the judge was more accurate than the median of the random comparison distribution. This procedure can be further refined by using only the estimations of the individual judge to generate a random comparison distribution instead of pooling the estimations of all judges (which was done with most measures in this study⁴).

Measures: Movie Case Test

Sound motion pictures were taken of a physician interviewing a patient. The patients were selected at

random from the medical wards of the Colorado General and Denver General Hospitals. Only those patients who were well enough to be moved to the room in which the filming took place, who had no language difficulty, and who were neither senile nor psychotic were selected. After the six-minute film was taken, the patients were administered the vocabulary section of the Wechsler-Bellevue Test, the Minnesota Multiphasic Inventory (card sort form), a reticence test, and a self-rating scale for a number of variables which were rough paraphrases of the MMPI Scales.

The task presented to the Ss was to estimate the patient's responses to the above measures after viewing the six-minute sound film. Thirty-six films were available, from which ten were selected at random. Each of the tests was scored by the Random-Comparison Method using the pooled estimations of all Ss to generate the difference scores as outlined below.

Reticence test. This measure concerned the reticence of the patients. The patients were asked to sort 48 cards containing statements about topics such as "your friends," "your cooperation with your doctor," "your attitude toward homosexuals," and so on. The patient was asked to sort these cards into two categories—those topics he was willing to discuss with the doctor in a subsequent interview in one category, those he was unwilling to discuss in another category. The entire list of topics and the actual instructions given to the patients were presented to the Ss in the test form. Their task was to estimate the number of topics that the patient was unwilling to discuss, i.e., the amount of the patient's reticence.

Vocabulary tests. The purpose of these measures was to discover how accurately the Ss could judge the vocabulary level of a patient from the film. The patients were administered the Vocabulary Section of the Wechsler-Bellevue Test. The list of words as well as the instructions given to the patients were presented to the students. They were asked to estimate the first word that the patient missed (Test A) and the three words in a row the patient missed (Test B).

Personality scales. The patients were administered the MMPI (9) and a series of 7-point self-rating scales that were paraphrases of the MMPI scales. The students were given two tasks: to estimate the point on each of the seven scales where the patient rated himself (Test A), and to estimate the point on the scales where the patient "really is" (Test B). The actual self-ratings served as criteria for the first task and the MMPI profile as a criterion for the second. The following 7-point scales were used by the patients for the self-ratings; the corresponding MMPI scales are in parentheses:

Scale 1 (Lie Score). How much are you concerned with what the doctor thinks of you as a person: extremely concerned, very concerned, more than most people, about the same as most people, less than most people, unconcerned, don't care at all?

Scale 2 (K Score). Do you feel that you are: extremely critical of yourself, overly critical of yourself, more critical of yourself than most people, just as critical as most people, less critical than most people, not very critical of yourself, never critical of yourself?

Scale 3 (Hypochondriasis). How do you feel about your health and your past illnesses? Have you had:

⁴ For a more detailed description of this procedure and a discussion of the types of test situations for which it is applicable, the reader is referred to (3, pp. 38-43).

more trouble than anyone you know, far more than most people, more than most people, about as much as most people, less than most people, far less than most people, less than anyone you know?

Scale 4 (Paranoia). When you have had dealings with people, how have you found them to be? Was there: no one you could really trust, a few you could trust, some you could trust, you could trust about half the people, more than half you could trust, almost all could be trusted, no one you couldn't trust?

Scale 5 (Psychasthenia). Some people are depressed, worry too much, have a lack of confidence, and are unable to concentrate. Would you say that you are usually: not at all like this, very little like this, some of it fits—most of it doesn't, some of it fits—some of it doesn't, most of it fits you, fits you well, fits you perfectly?

Scale 6 (Hypomania). How active a person have you been? Have you been engaged in: very many different activities, many activities, quite a few, some, not many, a few, very few different activities?

Scale 7 (Social Contact). Do you have your best times when you are: with a large number of strangers, with a large number of people—many of them strangers, with a large number of friends and strangers, with some friends and some strangers, with a small group of friends, with a few close friends, alone?

Measures: Sociometrics Test

The purpose of the Sociometrics Test was to measure the Ss' ability to estimate ratings in small groups. Ss were assigned at random to groups of ten members each, and were asked to place everyone in the group, including themselves, in rank order on the traits of leadership, cooperation, and likeability. This was called Task A. They were then asked to estimate where each individual would be ranked by the group as a whole for the same three traits (Task B). They were next asked to estimate how each of the nine others in the group had ranked them on these traits (Task C). And, finally, they were asked to estimate how each of the others in the group had ranked himself on the three traits (Task D). As criteria for the Ss' accuracy in each of the above tasks, the actual rank orders (Task A) were available. That is, Task A provided the group rank order, the way each person in the group ranked each S, and the way each person ranked himself.

Measures: Estimates of Group Opinion Test

All Ss answered a questionnaire concerning their attitudes toward medical education, and were then asked to estimate the percentage of the class who answered each question affirmatively. In all, 26 questions were used. Each student's accuracy was determined by correlating his estimated percentages with the actual percentage of the class who answered the corresponding questions affirmatively.

STUDY I

THE GENERALITY OF ACCURACY IN INTERPERSONAL PERCEPTION

The generality of accuracy of interpersonal perception over different procedures was studied by administering the measures de-

scribed above to a group of 65 Senior medical students. The Random-Comparison Method of scoring was used for the Ss' estimations of how each of the others in the group had ranked them (Task C) and for their estimations of how each of the others in the group had ranked himself (Task D). The estimations of all the Ss were not pooled as in the Movie Case test measures, but a separate random comparison distribution was computed for each S. This procedure was necessary because each S made estimations about the nine others in his group, who were therefore differently composed for each S. It was not necessary to employ the Random-Comparison Method with the estimation of the group rank order (Task B). Because the Ss were required to estimate a rank order, the estimations were necessarily spread evenly over the scale, and because the objects were assigned ranked scores, these scores were also evenly distributed across the scale. This procedure eliminates the effects of response set upon the Ss' accuracy scores. A difference score was computed between the estimated and actual group rank orders.

Results and Discussion

Table 1 presents the means and standard deviations of the accuracy scores. Apparently a sufficient range of ability was present to permit significant correlations among the ac-

TABLE 1
MEANS AND STANDARD DEVIATIONS OF ACCURACY SCORES*

Measure	Mean	SD
Reticence Test	44.77	12.2
Vocabulary		
Test A	68.92	19.8
Test B	72.62	14.5
Personality Scales		
Test A	65.85	6.1
Test B	62.55	9.1
Sociometrics		
Task B Lead.	19.63	6.8
Coop.	22.54	6.7
Like.	19.95	7.2
Task C Lead.	63.97	13.7
Coop.	65.89	12.9
Like.	67.91	13.7
Task D Lead.	67.43	12.1
Coop.	58.12	13.6
Like.	61.20	14.9
Group Opinion (Z-Score)	54.86	22.0

* N = 65.

TABLE 2
CORRELATION COEFFICIENTS AMONG MEASURES OF SOCIAL PERCEPTION ACCURACY*

Measure	Measure														Group Opin- ion Z- Score
	Vocabu- lary		Personality Scales		Sociometrics										
					Task B			Task C			Task D				
	A	B	A	B	Lead	Coop	Like	Lead	Coop	Like	Lead	Coop	Like		
Reticence Test	.19	.09	-.05	.02	-.06	-.09	-.33†	.07	-.21	.01	.13	-.08	.02	.09	
Vocabulary															
Test A		.19	-.13	-.00	-.15	-.11	-.18	.10	-.06	-.06	.25†	-.07	.25†	-.17	
Test B			-.07	-.07	.06	-.05	.06	.10	.05	-.04	.25†	-.13	.06	-.04	
Personality Scales															
Test A				.18	-.03	-.23†	-.04	.04	.11	.41†	.08	-.05	-.04	.06	
Test B					.06	-.09	.07	.14	-.08	.25†	.00	-.00	.09	-.05	
Sociometrics															
Task B Lead.						.15	.13	.01	-.08	-.19	-.10	-.11	.02	-.21	
Coop.							.39†	-.02	.00	-.12	-.06	.01	-.12	-.15	
Like.								-.21	-.16	.08	.05	-.04	-.19	-.19	
Task C Lead.									.08	.12	.19	.03	.01	-.03	
Coop.										.19	-.18	-.15	.04	.04	
Like.											.12	.13	-.05	.14	
Task D Lead.												.08	.02	-.07	
Coop.													.08	.11	
Like.														-.09	

* $N = 65$.

† Significant beyond the .05 level.

curacy scores, although the parameters of ability in interpersonal perception are as yet unknown.

Table 2 presents the product-moment correlation coefficients among the 15 measures of social perception accuracy. Only eight of these 105 correlations are significant beyond the .05 level of confidence, and two of these are negative relationships. These results are interpreted as failing to support the hypothesis of a general ability of interpersonal perceptiveness. Apparently interpersonal perceptiveness is more specific than previously assumed. Even though generality may be demonstrated with other measures and/or other Ss, these negative results make the *assumption* of generality untenable. Consequently, comparison of the results of interpersonal perception studies based on different measuring techniques can be justified only when their comparability has been empirically demonstrated.

STUDY II

THE GENERALITY-OVER-TIME OF ACCURACY AND RESPONSE SETS

The generality-over-time of accuracy and response sets was investigated by administering the Personality Scales test (see above)

three times, at six-month intervals, to the same group of Ss.

The Ss (not the same as in Study I) were 72 Senior medical students at the University of Colorado Medical School, who were tested at the beginning of their Senior year (Pretest), the middle of the year (Posttest-I), and at the end of the year (Posttest-II). The same measures were used, but the Ss made estimations for three different sets of ten patients.

Differential accuracy scores that were free of the influence of response sets were computed by use of the Random-Comparison Method as in Study I. In addition, three response-set characteristics were investigated in the present study:

1. The *implicit stereotype score*, defined as the mean of an S's estimations for all the objects on the same scale. For example, Ss were asked to estimate the self-ratings of ten patients on a seven-point scale "concern for what the doctor thinks of you as a person." The scale ran from "extremely concerned" to "don't care at all." An S may think that people *in general* are little concerned, or he may think that they are greatly concerned; this "stereotype" will determine to a certain degree the position on the scale around which

the perceived individual differences among the patients will cluster. We have used the mean of the estimations as an indicator of this stereotype.

2. The *adherence to stereotype score*, defined as the inverse of the average variance of an S's estimations for the scales. An S may perceive wide differences among people in the characteristic under consideration, or, he may perceive people as being very much alike. He may, in other words, cluster his estimations closely around his stereotype (mean estimation) or he may vary them widely. The variance of the estimations for each scale was computed and averaged for the seven scales as a measure of perceived individual differences.

3. The *assumed veridicality score*, defined as the inverse of the average difference between the S's estimation for the patient's self-rating and his estimation as to where the patient "really was." An S could assume a large measure of agreement between a patient's self-rating and "reality," or he could assume little agreement.

The Ss' stereotype, adherence to stereotype, and assumed veridicality scores were computed for each of the three administrations. The scores on each scale were combined to furnish composite scores for adherence to stereotype and assumed veridicality. The implicit stereotype scores were not combined but were treated separately since it would be expected that the stereotype would be unique for each scale. The scores on each administration were correlated with the scores on the other two administrations to determine the relationship between performance at one time and performance at the end of two successive six-month intervals.

Results and Discussion

Table 3 presents the product-moment correlation coefficients among the stereotype scores of the students on the Pretest, Posttest-I, and Posttest-II administrations. Of these 48 correlations only two failed to reach significance at the .01 level of confidence. Thus, the stereotypes that students hold regarding the patients on these scales were remarkably persistent over a six-month and one-year period.

Table 4 presents the product-moment correlation coefficients among the differential

TABLE 3
PRODUCT-MOMENT CORRELATION COEFFICIENTS
AMONG THE STEREOTYPE SCORES FOR THREE
ADMINISTRATIONS*

Measure	Pretest vs. Posttest I	Pretest vs. Posttest II	Posttest I vs. Posttest II
Task A			
Scale 1	.51	.54	.56
Scale 2	.51	.30	.36
Scale 3	.51	.45	.52
Scale 4	.58	.47	.41
Scale 5	.39	.54	.55
Scale 6	.59	.46	.66
Scale 7	.56	.47	.65
Scale 8	.52	.30	.47
Task B			
Scale 1	.44	.51	.63
Scale 2	.54	.68	.45
Scale 3	.45	.46	.54
Scale 4	.61	.36	.32
Scale 5	.57	.63	.65
Scale 6	.54	.54	.51
Scale 7	.67	.41	.43
Scale 8	.54	.44	.56

* $N = 72$ and $r = .31$ at the .01 level of confidence.

TABLE 4
PRODUCT-MOMENT CORRELATION COEFFICIENTS*
AMONG ACCURACY, ADHERENCE TO STEREOTYPE, AND
ASSUMED VERIDICALITY SCORES FOR THREE
ADMINISTRATIONS

Measure	Pre- Post I	Pre- Post II	Post I- Post II
Accuracy			
Task A	.19	.10	.24
Task B	.23	.28	.38
Adherence to Stereotype			
Task A	.47	.55	.66
Task B	.53	.50	.74
Assumed Veridicality	.64	.63	.76

* $N = 72$ and $r = .31$ at the .01 level of confidence.

accuracy, adherence to stereotype, and assumed veridicality scores for three administrations to the same Ss at six-month intervals.

Only one of the six correlations for the differential accuracy scores is significant at the .01 level of confidence and three are significant at the .05 level. While there seems to be a positive relationship among accuracy scores obtained at six-month and one-year intervals, the relationship is not a strong one as compared to the relationships indicated by the correlations for adherence to stereotype and assumed veridicality scores. All of these latter correlations are significant beyond the .01 level of confidence. The *largest* correlation for

accuracy was .38, while the *smallest* for the adherence to stereotype and assumed veridicality correlations was .47.

It should be remembered that while the format and content of the estimations which constituted the task for the *Ss* were identical, three different sets of films of ten patients each were administered. Furthermore, a full year of medical school activity intervened between performance on the Pretest and performance on Posttest-II, and six months intervened between the other administrations. The obtained correlations for stereotype, adherence to stereotype, and assumed veridicality are considered to be very high in view of these circumstances.

The relationship between performance at one time and performance six months and one year later that was found for the three response sets indicates that these characteristics are enduring and stable aspects of the *Ss'* approach to their task. The lower correlations that were obtained for the differential accuracy scores indicate that they are not as stable as response sets, in this situation at least. If it is true in general that, in the interpersonal perception task, response sets are more stable aspects of the *Ss'* performance than differential accuracy, then consistency in their performance may be due to consistency in response sets rather than to consistency in interpersonal perceptual ability. Since investigators of interpersonal perception have rarely eliminated response-set components from their accuracy scores, they are subject to precisely the above criticism. The reliabilities reported for accuracy of interpersonal perception should be seriously questioned because the consistency of the scores may be due to consistency in response set rather than to consistency in ability. Consequently, the authors maintain that differential accuracy of interpersonal perception as a reliable characteristic of *Ss'* performance *is yet to be established*.

Response sets deserve separate investigation aside from their influence upon accuracy scores. The *Ss* in this study had consistent modes of response to the judgmental task—they were consistent over time in using implicit stereotypes, consistent in the extent to which they adhered to these stereotypes, and consistent in the degree to which they as-

sumed veridicality between the patient's self-rating and "reality."

At present an investigation is being made of the personality correlates of such consistent response tendencies, and an attempt is being made to discover whether they have relevance for training in interpersonal relations.

SUMMARY

Two studies of the generality of interpersonal perceptiveness are reported. The first was an investigation of accuracy as a general ability over various procedures and the second was an investigation of the stability of accuracy scores and certain response-set scores over time. Differential accuracy was defined as the ability of an *S* to predict differences among other people (objects) on any item. In order to eliminate the influence of response sets upon the differential accuracy scores the Random-Comparison Method was used.

In the first study fifteen measures of interpersonal perception accuracy were administered to 65 Senior medical students. These measures were obtained from three procedures: (a) estimations of responses to personality inventories; (b) estimates of group opinion; and (c) estimation of ratings within small groups. These scores were intercorrelated to determine the generality of interpersonal perceptiveness, i.e., to determine whether or not subjects who were good (or poor) on one measure were good (or poor) on other measures. Only eight of the 105 correlations were significant at the .05 level of confidence and two of these were negative. The results were interpreted as failing to support the hypothesis of a general ability of interpersonal perceptiveness.

In the second study (involving a different group of *Ss*) the same measures of interpersonal perceptiveness were administered to 72 Senior medical students three times at six-month intervals. Thirty-six patients were selected at random (with certain restrictions) from the medical wards of two hospitals and were interviewed by a physician. Thirty-six, six-minute sound movies were taken of these interviews and ten of them were presented to the *Ss* at each of three test administrations. *Ss* were required to estimate the self-ratings of the patients and the point on the scale where the patient "really was" for seven per-

sonality scales. The criteria for accuracy were the actual self-ratings of the patients and the patients' performance on the MMPI.

The Random-Comparison Method was used to compute differential accuracy scores free of artifactual components. Three "response-set" scores were determined: (a) implicit stereotype score—the mean of an *S*'s estimations, (b) adherence to stereotype score—the inverse of the variance in his estimations, and (c) assumed veridicality score—the inverse of the difference between estimations for self-rating and for the patient's "real" score, summed for all seven scales and ten patients.

The scores on each administration were intercorrelated with similar scores on the other two administrations. The results indicated (a) significant but low relationships among accuracy scores over time, and (b) significant and high relationships among the response-set scores over time. These findings were interpreted to mean that response-sets were more stable elements of the *S*'s performance than differential accuracy and that previously reported reliabilities are spuriously high whenever the influence of response sets upon the accuracy scores has not been eliminated.

The conclusions from the two studies were as follows:

1. Failure to find support for the hypothesis of a general ability of interpersonal perceptiveness in this study makes the *assumption* of generality untenable. Consequently, comparison of the results of interpersonal perception studies based on different measuring techniques can be justified only when the comparability of such results has been empirically demonstrated.

2. The finding that response sets were more consistent over time than differential accuracy indicates that the reliabilities reported for

accuracy of interpersonal perception should be seriously questioned because the consistency of the scores may be due to consistency in response set rather than to consistency in ability. Since investigators have rarely eliminated response-set components from their accuracy scores, it is maintained that differential accuracy of interpersonal perception as a reliable characteristic of performance is yet to be established.

3. The high consistency of the response sets investigated in this study indicates that they deserve separate investigation aside from their influence upon accuracy scores.

REFERENCES

1. CAMPBELL, D. T. An error in some demonstrations of the superior social perceptiveness of leaders. *J. abnorm. soc. Psychol.*, 1955, **51**, 694-695.
2. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, **52**, 177-193.
3. CROW, W. J. A methodological study of social perceptiveness. Unpublished doctor's dissertation, Univer. of Colorado, 1954.
4. FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, **7**, 101-107.
5. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381-388.
6. GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411-422.
7. GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The intermediary key in the analysis of interpersonal perception. *Psychol. Bull.*, 1956, **53**, 258-266.
8. HASTORF, A. H., & BENDER, I. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 574-576.
9. HATHAWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI*. Minneapolis: Univer. of Minnesota Press, 1951.
10. TAFT, R. The ability to judge people. *Psychol. Bull.*, Vol. **52**, 1-23.

Received July 10, 1956.

PATTERNS OF ATTITUDES TOWARD CHILD REARING AND THE FAMILY

EARL S. SCHAEFER AND RICHARD Q. BELL

National Institute of Mental Health, National Institutes of Health, Public Health Service, U. S. Department of Health, Education, and Welfare

RESEARCH interest in the influence of the parent-child relationship upon personality development of the child has been stimulated by studies (4, 5, 8, 9, 11, 12, 17) which have reported that objectively measured attitudes of parents toward child rearing are related to emotional and social adjustment of children. Since it has been assumed that the behavior of parents with their children is a manifestation of personality functioning, Block's study (3), which demonstrated relationships between attitudes toward child rearing and personality characteristics of fathers, lends additional support to these findings. Clinical studies of mothers of maladjusted children, which have been reviewed by Symonds (18) and Radke (11), have also revealed syndromes of maternal behavior and attitudes which are related to differential personality structure of children. In the attitude studies cited, a number of parental concepts have been explored, but studies of a comprehensive sample of attitudes toward child rearing have not as yet been reported.

This study extends a series of investigations which have developed an objective inventory of attitudes toward child rearing (2, 14, 15) and a set of rating scales of maternal personality and of mother-child interaction (16). These research instruments are based upon an analysis of the domain of mother-child interaction in terms of concepts which are hypothetically related to the behavior of a mother with her child as it may affect the personality development of the child. Previous studies have demonstrated the feasibility of measuring the attitudes and behavior of women who are actually involved in relationships with their children. The present study was planned to determine whether it is possible to differentiate meaningful patterns of maternal attitudes in young unmarried women. Success in isolating theoretically significant patterns of maternal attitudes in this population would provide a basis for an attempt to predict mother-child interaction from attitude data gathered prior to the birth of a child.

METHOD

The Parental Attitude Research Instrument (2, 14), a questionnaire which had been developed to obtain multivariate measurement of attitudes toward child rearing and the family, was used to explore this domain. The 24 Likert-type attitude scales are listed in Table 1 with one sample item from each eight-item scale.

The subjects were 100 unmarried student nurses in their freshman and sophomore years. The homogeneity of the group in socioeconomic background, occupational goal, educational level, age, and intelligence permitted the control of a number of cultural factors to which variability in attitude might be attributed. Scores on the Otis Higher Intelligence Test which were available on a chance sub sample of 65 members revealed a median IQ of 111 with a range from 95 to 127. The inventory was given as an anonymous objective group test in regularly scheduled classes of psychology. Ss were asked to indicate whether they strongly agreed, mildly agreed, mildly disagreed, or strongly disagreed with each item. Scores for each scale were computed by summing the responses which had been converted into score weights of 1 through 4. The internal consistency reliability coefficients, computed with Kuder-Richardson Formula 20, are given in Table 2. A centroid factor analysis was performed on the product-moment interscale correlations, and the factors rotated obliquely to obtain simple structure (19).¹

RESULTS AND DISCUSSION

Despite the relatively low reliabilities of the scales, probably due to the homogeneity of the sample as well as the relatively short scales, the reliabilities were considered adequate for an investigation of the factorial structure of this domain. The rotated factor loadings on the five relatively independent factors which were isolated are given on Table 2. Scales which have their highest loading on a given factor are grouped under a tentative description or name of the factor which is given in italics. Since several scales have loadings above .30 on two factors, a

¹ The correlation matrix, the centroid factor matrix and the transformation matrix have been deposited with the American Documentation Institute. Order Document No. 5137 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1
NAMES OF SCALES WITH A SAMPLE ITEM FROM EACH

1. Encouraging Verbalization:
Children should be allowed to disagree with their parents if they feel their own ideas are better.
2. Infantilization:
A well-trained child always keeps the same close bond with its mother.
3. Seclusion of the Mother:
Too many women forget that a mother's place is in the home.
4. Breaking the Will:
Children need some of the natural meanness taken out of them.
5. Harsh Punishment:
If small children refuse to obey, parents should whip them for it.
6. Fostering Dependency:
A mother should do her best to avoid any disappointment for her child.
7. Martyrdom:
Children should realize how much parents have to give up for them.
8. Marital Conflict:
Sometimes it's necessary for a wife to tell off her husband in order to get her rights.
9. Strictness:
Strict discipline develops a fine strong character.
10. Irritability:
Children will get on any woman's nerves if she has to be with them all day.
11. Excluding Outside Influences:
Children should never learn things outside the home which make them doubt their parents' ideas.
12. Deification:
The child should be taught to revere his parents above all other grown-ups.
13. Suppression of Aggression:
There is no good excuse for a child hitting another child.
14. Rejection of the Homemaking Role:
Having to be with the children all the time gives a woman the feeling her wings have been clipped.
15. Autonomy of the Child:
Children should be given a chance to try out as many things on their own as possible.
16. Avoidance of Communication:
If you let children talk about their trouble they end up complaining even more.
17. Ignoring the Baby:
Handling a baby too much makes him spoiled and demanding.
18. Suppression of Sexuality:
It is very important that young boys and girls not be allowed to see each other completely undressed.
19. Intrusiveness:
An alert parent should try to learn all her child's thoughts.
20. Ascendance of the Mother:
A married woman knows that she will have to take the lead in family affairs.
21. Comradeship and Sharing:
If more parents would have fun with their children, the children would be more apt to take their advice.
22. Demand for Activity:
The sooner a child learns that a wasted minute is lost forever the better off he will be.
23. Abdication of the Parental Role:
Some children are so naturally headstrong that a parent can't really do much about them.
24. Acceleration:
A mother should make an effort to get her child toilet trained at the earliest possible time.

TABLE 2
ROTATED FACTOR MATRIX OF PARENTAL ATTITUDES

Factors with Scales Defining Each Factor	Reliability*	Factor Loadings				
		I	II	III	IV	V
I. Suppression and Interpersonal Distance						
21 Comradeship and Sharing	62	-.61	.06	.20	-.03	.06
15 Autonomy of the Child	61	-.57	.24	.20	-.24	.13
18 Suppression of Sexuality	58	.42	.01	.04	.40	-.05
16 Avoidance of Communication	59	.30	.09	.19	-.05	.22
1 Encouraging Verbalization	61	-.54	.51	.01	.03	-.16
II. Hostile Rejection of the Homemaking Role						
14 Rejection of the Homemaking Role	72	.01	.64	.17	.15	-.03
23 Abdication of the Parental Role	64	.00	.61	.00	.23	.00
10 Irritability	73	.07	.60	.04	-.19	.30
8 Marital Conflict	64	-.14	.47	.04	-.04	.17
17 Ignoring the Baby	77	.27	.38	.22	-.06	.03
III. Excessive Demand for Striving						
9 Strictness	74	-.07	.07	.58	-.19	.35
24 Acceleration of Development	64	.00	.05	.51	.10	.00
22 Demand for Activity	66	-.34	.26	.49	.11	.14
IV. Over-possessiveness						
7 Martyrdom	65	-.09	-.03	.36	.59	-.07
13 Suppression of Aggression	54	.00	.08	-.01	.57	.13
6 Fostering Dependency	60	.19	.06	-.14	.53	.13
19 Intrusiveness	84	.03	-.05	.12	.52	.01
2 Infantilization	57	.15	.18	.02	.49	.16
V. Harsh Punitive Control						
11 Excluding Outside Influences	67	.06	-.04	.41	-.02	.55
4 Breaking the Will	63	.04	.16	.26	-.01	.53
20 Ascendancy of the Mother	74	.44	.08	.00	.00	.52
12 Deification of the Parent	79	-.03	-.13	.41	.16	.45
3 Seclusiveness of the Mother	61	.23	-.16	-.05	.17	.42
5 Harsh Punishment	76	.13	.17	.15	-.27	.39

* Internal Consistency Reliability Coefficient—Kuder-Richardson Formula 20.

scale may be mentioned twice in the description of factors below.

The factor of Suppression and Interpersonal Distance is defined by both negative and positive loadings. The attitudes of Comradeship and Sharing, Autonomy of the Child, and Encouragement of Verbalization seem to reflect approval of a close relationship with a child which still permits him freedom to develop as a differentiated individual and also to express his views about family affairs and parent behavior toward him. The other pole of this factor reflects an attitude that the mother should dominate the family, that the sexual interests of the child should be tabooed, that children should not be allowed to talk about their fears and hostilities with their parents, and that a baby needs only physical care without social play and tenderness from the mother. This factor seems to be related to the warmth and closeness of the relationship and

the degree of tolerance for expression of individuality of the child.

The content of the factor, Hostile Rejection of the Homemaking Role, reveals dissatisfaction with the role of housewife and mother and a high level of hostility which is openly expressed. The scale, Rejection of the Homemaking Role, consists of items which state the unhappiness of a woman at her loss of freedom and her dissatisfaction with the duties of caring for a home and children. An attitude that one can do very little in rearing children, with an implied denial of the responsibility of a parent, is revealed in Abdication of the Parental Role. Irritability items imply impatience, intolerance, and a low threshold for the arousal of hostile reactions toward children. In the scale Encouraging Verbalization the *S* reveals her approval of a child's open expression of his disagreements with and hostility toward his parents. Marital Conflict reflects

an expectation of a great deal of quarreling and fighting between husband and wife and again reflects hostility which is openly expressed. Items on Ignoring the Baby express the opinion that a baby needs only physical care and that much of the attention given to him is unnecessary. This factor may be related to the pattern of maternal rejection which has been described by Newell (10), Symonds (18) and others.

The factor, Excessive Demand for Striving, is defined by scales which reveal a demand for conformity and achievement by the child to meet the parents' needs. Strictness demands conformity to rigid rules; Acceleration of Development states that a child should reach developmental goals at an accelerated pace; and Demand for Activity items reflect a belief that a child should be kept busy and should achieve as much as possible. Exclusion of Outside Influences, Deification of the Parent, Martyrdom, and Breaking the Will may be related to methods of enforcing these demands for conformity and achievement. This factor to some degree resembles Lorr and Jenkins' (7) factor, "organization and effectiveness of control in the home," derived from a second order factor analysis of the Fels Parent Behavior Rating Scales.

The factor of Over-Possessiveness seems to have a central theme of control of the child through keeping the child indebted to the mother, dependent, and immature. The highest loading, on Martyrdom, reveals a perception of a mother as a sacrificing, suffering woman with an implied obligation of the child to feel guilty and indebted to her. Suppression of Aggression indicates a wish for a quiet child who would be more easily controlled than an active, aggressive one. Infantilization refers to a preference for an immature child who is closely tied to his mother's apron strings, and Fostering Dependency reflects a perception of the child as a weak person who must be protected by the mother. Suppression of Sexuality indicates a taboo on sexual curiosity as well as behavior. A belief that the mother should know everything the child is thinking and doing is revealed by the Intrusiveness scale. This factor resembles to some degree the factor of maternal behavior described by Roff (13) as "concern for the child," Lorr and Jenkins' (7)

factor of "dependency encouraging," and Levy's (6) concept of "maternal over-protection."

The factor of Harsh Punitive Control resembles the concept of parental dominance which has been described by Symonds (18) among others. In this factor the scale, Exclusion of Outside Influences, reveals a wish for absolute control by the parents which excludes any other source of influence. Enforcement of parental authority through fear and forcing the child into submission is revealed in Breaking the Will. A need to dominate the entire family is shown by the assumption that the mother should direct all family activities. This factor also contains a scale which states that the mother should avoid an active social life, and another that approves of the use of harsh physical punishment as a method of influencing the child's behavior. The loading on Irritability reflects the extent to which the respondent feels that rearing children can be an irritating burden; and Strictness reveals an opinion that rules should be rigidly enforced. This factor of Hostile Punitive Control seems to resemble the negative pole of Roff's (13) factor of "democratic guidance" and of Lorr and Jenkins' (7) second factor of "democracy and democratic values."

Several of the factors appear to correspond to parent-child interaction patterns as described by clinicians, and to factors that have been isolated from the Fels parent behavior ratings (7, 13). The apparent similarity of these patterns arrived at from different methods of studying parent-child relationships suggests the need for studies of the degree of correlation between measures of attitudes toward child rearing and observations of actual behavior with children. The relationship between these attitudes toward child rearing and other personality variables of the respondents should also be investigated. This study, in conjunction with previous studies of objectively measured attitudes toward child rearing, suggests that prediction of mother-child interaction and of personality development of children may be possible since measurable variability in attitude and an organized structure of attitudes toward child rearing can be determined in young unmarried women.

SUMMARY

Twenty-four scales of the Parental Attitude Research Instrument were administered to 100 student nurses. Internal consistency reliability coefficients were found which permitted a centroid factor analysis of the scales. An oblique simple structure rotation revealed five relatively independent factors, identified as Suppression and Interpersonal Distance, Hostile Rejection of the Homemaking Role, Excessive Demand for Striving, Over-possessiveness, and Hostile Punitive Control. The resemblance of these attitude factors to syndromes of parental behavior and attitudes as described by clinicians in parents of emotionally disturbed children and to factors that have been isolated from the Fels Parent Behavior Scales was noted. It is suggested that this measuring instrument may permit prediction of different patterns of mother-child interaction.

REFERENCES

1. BALDWIN, A. L., KALHORN, JOAN, & BREESE, FAY. Patterns of parent behavior. *Psychol. Monogr.*, 1945, **58**, No. 3 (Whole No. 268).
2. BELL, R. Q., & SCHAEFER, E. S. Scales for measuring general orientations of parents toward children. *Amer. Psychologist*, 1955, **10**, 439. (Abstract)
3. BLOCK, J. Personality characteristics associated with fathers' attitudes toward child-rearing. *Child Develpm.*, 1955, **26**, 41-48.
4. FREEMAN, R. V., & GRAYSON, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **50**, 45-52.
5. HARRIS, D. B., GOUGH, H. G., & MARTIN, W. E. Children's ethnic attitudes: II. Relationship to parental beliefs concerning child training. *Child Develpm.*, 1950, **21**, 169-181.
6. LEVY, D. M. *Maternal over-protection*. New York: Columbia Univer. Press, 1943.
7. LORR, M., & JENKINS, R. L. Three factors in parent behavior. *J. consult. Psychol.*, 1953, **17**, 306-308.
8. MARK, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 185-189.
9. MILES, KATHERINE A. Relationship between certain factors in the home background and the quality of leadership shown by children. Unpublished doctor's dissertation, Univer. of Minnesota, 1945.
10. NEWELL, A. W. The psychodynamics of maternal rejection. *Am. J. Orthopsychiat.*, 1934, **4**, 387-401.
11. RADKE, MARIAN J. Relationship of parental authority to children's behavior and attitudes. *Univ. Minn. Inst. Child Welf. Monogr.*, 1946, No. 22.
12. READ, KATHERINE H. Parents' expressed attitudes and children's behavior. *J. consult. Psychol.*, 1945, **9**, 95-100.
13. ROFF, M. A factorial study of the Fels Parent Behavior Rating Scales. *Child Develpm.*, 1949, **20**, 29-44.
14. SCHAEFER, E. S., & BELL, R. Q. Structure of attitudes toward child-rearing and the family. *Amer. Psychologist*, 1955, **10**, 319-320. (Abstract)
15. SCHAEFER, E. S., & BELL, R. Q. Parental attitude research instrument (PARI). Normative data. Unpublished manuscript. Library, National Institute of Mental Health, Bethesda, Md., 1955.
16. SCHAEFER, E. S., BELL, R. Q., & BAYLEY, NANCY. Quantification of maternal behavior and consistency of mother-child interaction. *Amer. Psychologist*, 1956, **11**, 404. (Abstract)
17. SHOEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, **39**, 101-148.
18. SYMONDS, P. M. *The psychology of parent-child relationships*. New York: Appleton-Century Co., 1939.
19. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.

Received July 20, 1956.

AUTHORITARIAN IDEOLOGY IN NEGRO COLLEGE STUDENTS¹

GEORGE A. STECKLER²

Cleveland Receiving Hospital and State Institute of Psychiatry

Most of the research stemming from *The Authoritarian Personality* (2) has concerned members of the majority group in America. Recent studies have explored ideology in the Jewish minority (1, 11, 12). The present study is an attempt to clarify the effects of minority group membership on the ideology of Negroes.

The psychological situation of the average American Negro is both ambiguous and conflict ridden. According to frustration-aggression theory (6), discrimination should arouse the Negro's aggression toward the majority group. However, the Negro may be fearful of retaliative action by the powerful majority, and he also may identify with the values of the majority, especially if he is of the middle class. He may therefore be expected to have ambivalent feelings regarding the majority group as well as his own. What are the ideological manifestations of this dilemma?

METHOD

Procedure and Sampling

The questionnaire method was used. The research was divided into two phases. In the first phase, preliminary forms of opinion-attitude scales were administered to 299 Ss, mostly members of church groups and students attending predominantly Negro colleges. The purpose of the preliminary administration was to ascertain reliability and internal consistency for the scales. On the basis of obtained data, the scales were revised and shortened.

For the study proper, Ss were obtained from colleges and universities attended primarily by Negroes. Cooperation was sought by letter from nine psychology instructors, all members of APA. Seven of the nine instructors responded and subsequently administered questionnaires to their classes.³ Questionnaires were filled out anonymously. Approximately 85 per cent of the questionnaires mailed out were completed and returned. The schools from which the Ss were obtained

were diverse in both location and character. North, South, and border states were represented. A teachers' college, an agricultural and mechanical school, and liberal arts colleges were included. Mean age of Ss was 20.5 years. Nearly half were Southerners; the remainder were Northerners or had lived for extended periods in both North and South. Although the group was predominantly working class in origin, family socioeconomic status was considerably higher than for the general population of Negroes.

Variables and Scales

Variables chosen for the study were: authoritarianism (F), ethnocentrism (E), politico-economic conservatism (PEC), anti-Negro ideology (AN), and anti-white ideology (AW). The California F, E, and PEC scales (2) were used to measure the first 3 of these variables. Separate scales were constructed to measure anti-Negro and anti-white ideology. Anti-Negro items were therefore omitted from the E scale. All scales were of the seven-point Likert type.

Anti-Negro and anti-white ideologies were conceived of as relatively organized and stable systems of opinions, values, and attitudes regarding Negroes and whites. They involve negative, stereotyped opinions and hostile attitudes regarding Negroes and whites. They also include moral values which justify the attitudes and permeate the opinions. In order to gain a broad understanding of anti-white and anti-Negro ideology as it exists among Negroes, Negro newspapers, magazines, and literature were perused. In formulating the items for the anti-Negro (AN) scale, it was assumed that virulent, extreme anti-Negro ideology is not common among Negroes themselves. In order to make the scale discriminating as well as nonrepugnant to the Ss, the items were made milder than anti-Negro ideology current among whites. However, the content of this current ideology was retained. Sixteen of the most discriminating items were selected from the preliminary form of the scale. AN items are presented in Table 1. All are phrased in the anti-Negro direction. Most of the items stress the offensiveness of Negroes and the opinion that their deviant behavior causes prejudice.

The revised anti-white (AW) scale contains 18 of the most discriminating items from the preliminary form. All but one are phrased in the anti-white direction. The items assert that whites are intrusive, untrustworthy, dishonest, and inferior in certain spheres, and advocate that they be deprived of their present power. The AW scale is presented in Table 2.

In order to shorten the length of the questionnaire and therefore economize on classroom time, two forms of the questionnaire were composed. Form A contained AN, AW, and F. Form B substituted E and PEC in place of F.⁴

¹ This article is based in part on a doctoral dissertation submitted to the Department of Psychology, Western Reserve University, February 1954. The author wishes to thank Drs. William R. Morrow and Richard W. Wallen who served as chairmen of his doctoral committee.

² Now at Cleveland Regional Treatment Center.

³ The following individuals graciously consented to help: Profs. J. Otis Smith, Vera D. Hunton, Regina M. Goff, Leonora C. Lane, Paul Clifford, James F. Condell, and James T. Taylor.

⁴ The difference between the means on the two scales (AN and AW) shared in common by both forms was not significant.

TABLE 1
ANTI-NEGRO SCALE ITEMS*

1. A large part of the problems facing Negroes today are caused by Negroes themselves.
2. I would rather not marry a person who has very kinky hair, wide nostrils, and thick lips.
3. The lower-class Negro is to blame for a lot of anti-Negro prejudice.
4. Whites and Negroes can get along on jobs until too many Negroes try to push themselves in.
5. One big reason why racial prejudice is still so strong is that Negroes offend people by being so sensitive about racial matters.
6. One important reason why Negroes are discriminated against in housing is that they don't keep up the property.
7. One reason why racial prejudice still exists today is the fact that many Negroes are dirty, loud, and generally offensive in their ways.
8. One trouble with Negroes is that they are even more jealous of each other's success than are whites.
9. Too many Negroes have abused the privilege of attending baseball games by being rowdy, noisy, and cheering only for the colored ballplayers.
10. Segregation and jimcrow will never end unless the average colored person becomes better educated and better mannered.
11. Colored people can hardly be expected to gain social equality until many more of them exert some effort to better themselves and live more decently.
12. With all of the drinking, cutting, and other immoral acts of some Negroes, white people are almost justified for being prejudiced.
13. Too many Negroes, when they get a little money, spend it all on whiskey, flashy cars, or expensive clothes.
14. One is almost ashamed to be a Negro when he sees so many of them who look and act like cotton pickers fresh from the fields.
15. A great many Negroes become officious, overbearing, and disagreeable when given positions of responsibility and authority.
16. Negroes would solve many of their social problems if so many of them were not irresponsible, lazy, and ignorant.

* Several items are adapted from Pitt (10).

RESULTS

Odd-even reliability coefficients, corrected by the Spearman-Brown prophecy formula, were computed for each scale. Table 3 presents the reliability and related statistics.

Four of the scales yielded high reliability coefficients, indicating relative freedom from chance errors of measurement and comparative stability of scores. Low reliability of PEC apparently is not solely a function of its shortness (10 items). The California study obtained much higher reliabilities for their somewhat longer version of this scale.⁵ Morrow (9) re-

⁵ By means of the Spearman-Brown prophecy formula, it was estimated that increasing the present scale to 16 items would probably increase reliability to

TABLE 2
ANTI-WHITE SCALE ITEMS*

1. There is nothing lower than white trash.
2. White people may be all right, but they carry it too far when they try to butt into the Negro's affairs and go around with Negro women.
3. The whites have shown by their actions that they are naturally immoral, vicious, and untrustworthy.
4. No matter how nicely they treat a colored person, white people don't really mean it.
5. It is usually a mistake to trust a white person.
6. Any Negro who marries a white is a traitor to his people.
7. There may be a few exceptions, but white musicians and athletes are definitely inferior to Negro musicians and athletes.
8. White people are only friendly to Negroes when they want something out of them.
9. Negroes can expect no real help from white people in the fight against racial discrimination.
10. Most white people are always looking for ways to cheat and steal from the colored people.
11. The colored race has been pushed around long enough; it's about time that the whites were made to get out of the Negro communities.
12. If there is a Heaven, it is hard to imagine that there are many white people up there.
13. Although the white man now rules the world, it will be a happy day when the tables are turned and the colored people become the rulers.
14. The world might be a better place if there were fewer white people.
15. When the Bible says, "The bottom shall rise to the top,"† it gives hope that the Negro people will someday give the orders in this country instead of whites.
16. It may be wrong to damn all whites, but it's plain that whites have all the money and power, and that they look down on anyone who is colored.
17. There are many white people who are not prejudiced and who sincerely believe that Negroes are equals.‡
18. When it comes to such things as sports, dancing, music, and making love, the white man is not as talented as the Negro.

* Several items are adapted from Pitt (10).

† The Bible quotation is inaccurate.

‡ This item is positively stated: the remaining ones are negative.

TABLE 3
RELIABILITY AND RELATED STATISTICS FOR THE SCALES

Property	Scale				
	E	AN	AW	F	PEC
Reliability	.75	.84	.88	.81	.44
Mean	3.58	4.55	2.75	4.66	4.39
SD	.99	1.06	1.12	.93	.75
Range	1.2-5.8	1.1-6.6	1.0-6.6	1.4-6.6	1.5-6.0
N	212	449	449	222	210
No. of Items	12	16	18	20	10

ports high reliability in studies using the present PEC scale with white college students.

.56, well below the r of .73 obtained by the California study for its 16 item PEC scale.

TABLE 4
INTERCORRELATIONS AMONG THE SCALES*

Scale	AN	AW	E	F
AN		.24	.40	.57
AW			.60	.55
E†				...

* All correlations are significant beyond the .01 level.

† The correlation between E and F was not computed because each scale was in a different form of the questionnaire.

Frequency distributions approached normality for all scales with the exception AW, which was markedly weighted with low scores and skewed in the positive direction. However, the range of scores on AW is wide and variability is greater than for any of the other scales. In spite of the bunching of low scores, the scale evidently offered sufficient opportunity for freedom of expression, agreement or disagreement.

Taking a mean score-per-item of 4.0 as the theoretical neutral point for each scale, it is apparent that the sample tends to agree with anti-Negro, authoritarian, and conservative ideologies and to disagree with ethnocentric and anti-white ideologies. The obtained means are similar in degree to those obtained in the preliminary phase of the present study. The present group's standing on AN and AW coincides with the findings of both Pitt (10) and Bayton (4), whose Negro groups are ideologically similar in spite of differences in socioeconomic class. It should be noted that only one group studied by the California investigators—prison inmates—scored as high as the present sample on the F scale. However, our sample's E scale mean of 3.58 is lower than that of many of the California samples.

The correlation coefficients listed in Table 4 indicate slight to substantial relationships between scores on the several scales. This might be anticipated because each of the four scales is a measure of authoritarian ideology.

DISCUSSION

In view of the interrelationships between scales and the patterning of scale means, it is suggested that a more or less consistent psychological trend exists in the Negro middle-class. There is no clear evidence of a blind "response set to acquiesce" which Bass (3) believes can account for many correlations of the F scale; most Ss, even many high on F, did not "ac-

quiesce" to the negatively formulated AW scale. The evidence does indicate that a particular kind of authoritarianism—identification with and conformity to values and norms of the white majority—is prevalent among middle-class Negroes. Like authoritarian Jews (1), they are prone to view prejudice as caused by the deviant behavior of certain members of the minority, and they are loath to make generalized criticisms of the majority group. They readily agree with the phenomenally conventional, moralistic, and ultraconservative F scale items, and their politico-economic attitudes are largely conservative.⁶ It is suggested that the standards with which Negro students identify are the narrowly perceived and rigidly adhered to values of the white middle class. This is not to say that the Negro middle class is altogether antidemocratic; the sample tended to reject ethnocentric items, perhaps because they have become sensitized to overt prejudice and because Negro authorities strive to foster tolerance (5).

One interesting aspect of the results concerns those Ss who score high on the F scale. These Ss not only accept most of the anti-Negro stereotypes, but they also agree with many of the anti-white statements. Their ideological position is somewhat contradictory in that they identify with white middle-class standards, yet are critical of white people. However, these Ss are Negro despite their conflicts about it, and whites do not accept them as individuals. They can agree with AW items because these refer primarily to specific behaviors of whites rather than to white middle-class ideology. They implicitly deny any connection between the behavior and the ideology.

These findings may be understood, perhaps, as a partial "identification with the aggressor" (7). By accepting the respectability and propriety of white middle-class ideology and by dissociating himself from Negroes in general the Negro college student apparently attempts to deny the feeling (8) that he himself may be unworthy of acceptance by the majority.

⁶ No direct evidence is available as to why the Ss are inconsistent in their politico-economic conservatism but we may speculate as follows: Although essentially conservative, middle-class Negroes do have an interest in the adoption of some liberal measures as a means of increasing their status. Thus, their politico-economic attitudes, though inconsistent from the standpoint of whites, are actually consistent with their general outlook.

SUMMARY

The aim of this study was to investigate authoritarian ideologies in a group of Negro Ss. The questionnaire method was used. Opinion-attitude scales of the Likert type were constructed to measure anti-Negro and anti-white ideologies. The California E scale (anti-Negro subscale omitted), F scale, and a revision of the California PEC scale were also used. Questionnaires were administered to a large sample of students attending colleges with primarily Negro enrollment.

Results were as follows:

1. Intercorrelations of the scales were positive and, for the most part, of moderate degree.

2. Mean scale scores indicated group tendencies to disagree with ethnocentric and anti-white ideologies, but to agree with authoritarian, anti-Negro, and conservative ideologies.

The data were interpreted to indicate an attempt by the Negro middle class to identify with stereotyped white middle-class values and to dissociate itself from other Negroes.

REFERENCES

1. ADELSON, J. B. A study of minority group authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 477-485.

2. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
3. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 613-623.
4. BAYTON, J. A. The racial stereotypes of Negro college students. *J. abnorm. soc. Psychol.*, 1941, **36**, 97-102.
5. BIRD, G., MONACHESI, E. D., & BURDICK, H. Infiltration and the attitudes of white and Negro parents and children. *J. abnorm. soc. Psychol.*, 1952, **47**, 688-699.
6. DOLLARD, J., DOOB, L., & MILLER, N. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.
7. FREUD, ANNA. *The ego and the mechanisms of defence*. New York: International Universities Press, 1946.
8. KARDINER, A., & OVESEY, L. *The mark of oppression*. New York: Norton, 1951.
9. MORROW, W. R. Personal communication. 1953.
10. PITT, W. A study of Negro ideological tendencies. Unpublished master's dissertation, Western Reserve Univer., 1951.
11. RADKE-YARROW, MARIAN, & LANDE, B. Personality correlates of differential reactions to minority group belonging. *J. soc. Psychol.*, 1953, **38**, 253-272.
12. SARNOFF, I. Identification with the aggressor: some personality correlates of anti-semitism among Jews. *J. Pers.*, 1951, **20**, 199-218.

Received July 20, 1956.

PERSONALITY, ANXIETY, AND MUSCLE TENSION AS DETERMINANTS OF RESPONSE STEREOTYPY

JAMES L. KUETHE AND CHARLES W. ERIKSEN¹

The Johns Hopkins University

EXPERIMENTAL investigations of the effects of anxiety upon performance have succeeded in showing a number of ways that anxiety influences behavior. Little research emphasis, however, has been devoted to determining whether the observed performance effects are due to anxiety *per se* or whether they are characteristic of any strong drive state. Similarly, there has been little effort to determine which of the many properties of anxiety is responsible for the observed changes in psychological functioning.

Anxiety as clinically observed and defined has a number of different properties. Not only are there objectively observable autonomic nervous system indicators, but one of the most prominent symptoms of anxiety is the increase in generalized muscle tenseness or activity. In popular terminology the anxious person is "tense." Meyer (9), and Matarazzo, Ulett, and Saslow (8) have suggested that certain of the effects of anxiety upon responses may be due to the accompanying muscular tension. The plausibility of this suggestion is supported by many striking parallels between current research on anxiety effects and the older research on the effects of experimentally induced muscular tension so well summarized by Courts (2).

The purpose of the present study was to determine whether certain of the effects of anxiety upon performance could be traced to the increased muscular tension that accompanies anxiety. To accomplish this end, the stereotypy and bias with which Ss used a series of responses under conditions of experimentally induced muscle tension and experimentally induced anxiety were studied.

Design

It has been previously shown (5) that increased bias and stereotypy occur in the use of number responses on an absolute judgment task when the Ss are made anxious through random electric shocks. In the present study, a similar response effect was employed, but in order to increase the opportunity for bias to occur, the responses were freed from external

stimulus determination. This was accomplished by disguising the experiment as an extrasensory perception study. Ss were told that *E* would concentrate on the numbers 1-11, one at a time, and they were to try to determine through ESP what number *E* was looking at. The frequency and the sequential order with which the Ss used the different number responses was recorded.

There were three experimental conditions. In the anxiety condition, Ss were randomly shocked on the ankle during the course of the experiment. In the muscle tension condition, Ss were required to squeeze a hand dynamometer with a predetermined force prior to and during each response. The third condition was a control where neither anxiety nor muscle tension was experimentally introduced.

As a further test of the hypothesis that the anxiety effects on this task are due to the accompanying muscular tension, Ss were selected on the basis of high or low scores on the psychasthenia scale from the MMPI. Previous research (3, 4) has shown that there is an interaction between high and low psychasthenia scores and performance under conditions of induced anxiety. If a similar interaction between psychasthenia extremes and muscle tension were found to occur, then the hypothesis that anxiety influences performance through increased muscle tension would be more strongly substantiated.

METHOD

All Ss were run individually. At the beginning of the experiment they were given a brief summary of the studies that have been done on ESP and were told that the purpose of the present experiment was to duplicate some of the findings. The *E* announced that he would draw a number from 1-11 from a pack of cards and concentrate on it; when the signal light flashed, *S* was to give the first number in this range that came to his mind. The Ss were shown a large pack containing numbered cards which *E* supposedly would draw from. The Ss were informed that the numbers 1-11 would occur in a random order, and care was taken to insure that the Ss understood the concept of randomness. The Ss assigned to the control condition were given no further instructions.

The Ss assigned to the muscle tension condition were further instructed that *E* was interested in the effects of muscular activity on the ESP phenomena.

¹ Now at the University of Illinois.

Subjects in this group were asked to squeeze a hand dynamometer with maximum force, using their preferred hand. One third of this force was the level of tension *S* had to maintain while responding in the experiment.

In the anxiety condition, *Ss* were told that *E* was interested in the effect of electrical stimulation upon ESP performance. They were fitted with ankle electrodes through which faradic shocks were given by means of an inductorium supplied with a 3-volt D.C. source. The *Ss* were carefully instructed that the shocks they would receive were in no way connected with their responses or their performance on the task. Prior to beginning the experiment, the maximum shock intensity that each *S* would tolerate was determined.

All *Ss*, regardless of experimental condition, received a response signal at 15-sec. intervals and were given a minute rest period at the end of each series of 12 responses. The frequent rest periods, while the same for all *Ss*, were primarily for the benefit of the muscle tension group in order to avoid their becoming fatigued. During the experiment each *S* responded a total of 154 times.

In the muscle tension group a special signal light flashed five seconds before the appearance of the response signal. When the special light flashed, *Ss* were required to squeeze the dynamometer with the predetermined force (one-third of their maximum effort) and to maintain this grip until the special light went off five seconds after the response signal had been given.

The anxiety group received 25 shocks randomly distributed through the 154 trials. Seven of these shocks were of maximum intensity for the particular *S*, 10 medium, and 8 weak. No shocks were given until 10 seconds had elapsed since the last response in order to disconnect the two events as much as possible. The same random shocking order was used for all *Ss* in this group.

Apparatus

The *Ss* sat on the opposite side of a desk from *E*, separated by a wooden partition mounted on the desk. The partition contained two translucent windows which could be illuminated by *E*. A flash of light in one of the windows was the signal for *S* to respond. The other window was used only for the muscle tension group to signal when to grip the dynamometer. In order to insure that *Ss* in the muscle tension group were maintaining the required tension, the dynamometer was mounted on the desk and could be observed by *E* through an aperture in the partition.

Subjects

The *Ss* were male undergraduate students from the introductory psychology course at the Johns Hopkins University. They were required to serve a minimum number of hours as experimental *Ss* as part of the course requirements. The group form of the MMPI was administered to the entire class of 154 students, and the *Ss* used in the present experiment were obtained from those students scoring in the highest and lowest quartiles of the class on the psychasthenia scale. In each of the three experimental conditions there were 16 *Ss*,

eight whose scores were in the upper quartile on psychasthenia and 8 whose scores fell in the lower quartile.

RESULTS

In order to measure the stereotypy or predictability with which *Ss* used the eleven response numbers, uncertainty statistics derived from information theory (7) were used. The basic measure, response uncertainty (U_r), is equal to $\sum p \log_2 p$, where p is the proportion of the time each response occurred. In the present study where *Ss* had available 11 different responses, response uncertainty would have had a maximum value of 3.48 bits if *S* had used each of the 11 responses equally often. The extent to which *S* departed from equal usage of all of the 11 responses is reflected in a corresponding reduction in the value of U_r . It is to be noted that failure to use all responses equally often decreases uncertainty as to what an *S*'s response will be or—alternatively stated—it increases the predictability of his behavior.

The value of U_r was computed for each *S* in each of the six subgroups. These values were then subjected to a two-way classification analysis of variance (psychasthenia groups and experimental conditions). The only significant effect obtained from this analysis was the interaction of the high and low psychasthenia groups with the three experimental conditions. ($F = 3.44$, $p < .05$ with df of 2 and 42). The nature of this interaction can be seen in Table 1 where the mean U_r is shown for each of the subgroups.

The table shows that *Ss* with low scores on psychasthenia are less predictable or biased in their use of the eleven response numbers under the control condition than are the *Ss* with high psychasthenia scores. However, when anxiety or muscle tension is introduced, the low group

TABLE 1
AVERAGE RESPONSE UNCERTAINTY, U_r , IN BITS AS
A FUNCTION OF PSYCHASTHENIA AND
EXPERIMENTAL CONDITIONS

	Low Psychasthenia Group	High Psychasthenia Group
Control	3.41	3.26
Muscular Tension	3.31	3.37
Anxiety	3.30	3.33

Using the estimate of the population variance, .0143, obtained from the error term in the analysis of variance, mean differences of .12 and .16 are significant at the .05 and the .01 levels of confidence, respectively.

becomes more predictable or biased while the high group becomes less so. If the individual cell means are compared, it is found that the mean difference between the high and the low groups in the control condition is significant beyond the .02 level of confidence, but the difference between the control and the anxiety and muscle tension conditions within the high and low groups only reach significance levels of about .10. If, however, the tension and anxiety conditions are combined in the high and low groups, respectively, and tested against the corresponding control condition, it is found that the high psychasthenia *Ss* show a significant decrease in predictability or bias of responses under tension and anxiety ($p < .05$) while the low psychasthenia *Ss* show a significant increase ($p < .05$). Such a combination of groups is justified since the mean differences between the anxiety and tension conditions do not begin to approach acceptable significance levels in either the high or low psychasthenic group.

This analysis of response uncertainty, U_r , deals with one way in which an *S's* responses are predictable. It reveals the extent to which *S* demonstrates number preferences or biases in the use of the 11 response numbers. There is, however, another dimension along which an *S's* behavior can be described as stereotyped or predictable. This is in terms of the second order or sequential dependencies among his responses. For example, an *S* may use each of the 11 response numbers equally often but persist in following his response "2" with the response "8" and response "5" with response "9." In this case, his behavior would show a fairly high degree of predictability since if one knew what his preceding response had been, one would be able to make a fair prediction of his next response.

To determine and compare the sequential dependencies in the present data, uncertainty

TABLE 2
AVERAGE CONDITIONAL UNCERTAINTY, $U_{r-1,r}$, IN BITS
AS A FUNCTION OF PSYCHASTHENIA AND
EXPERIMENTAL CONDITIONS

Conditions	Groups	
	Low Psychasthenia	High Psychasthenia
Control	2.73	2.58
Muscular Tension	2.65	2.71
Anxiety	2.60	2.65

statistics were again employed. The conditional uncertainty measure, $U_{r-1,r}$ gives the uncertainty in bits in predicting a response, r , when the preceding response, $r-1$, is known (7). Mathematically, $U_{r-1,r}$ is defined as $\sum p_1 \sum p_2 \log_2 p_2$, where p_1 is the probability of any given response occurring and p_2 is the probability for each response following a specified response. As in the previous analysis, maximum uncertainty would equal 3.48 bits if there were no sequential dependencies in the responses and each of the 11 responses was used equally often.

The conditional uncertainty in each *S's* responses was computed and analyzed via a two-way classification covariance analysis (psychasthenia groups and experimental conditions). Since the value of U_r sets an upper limit upon the value that $U_{r-1,r}$ can assume, it was necessary to make a correction via the covariance method for mean differences between the subgroups on U_r . The results of this covariance analysis were essentially similar to those obtained from the analysis of the response uncertainty measures. Neither of the main effects was significant, but again there was a significant interaction between the high and low psychasthenia groups and the three experimental conditions ($F = 3.86$, $p < .05$). The nature of this interaction can be seen in Table 2 where the unadjusted mean conditional uncertainties are shown for each of the six subgroups.

As is evident, the pattern of the interaction is the same as that obtained in the previous analysis of U_r . The low psychasthenia group shows less stereotypy in the control condition, but under tension or anxiety these *Ss* show an increase in stereotypy or response predictability. The high psychasthenia *Ss* show the reverse effect. In the control condition they are more stereotyped than the low *Ss*, but under tension or anxiety their behavior becomes less stereotyped or predictable. Again it is to be noted that there is little difference between the anxiety and muscle tension conditions in either the high or low psychasthenia group.

DISCUSSION

The results are consistent with the hypothesis that the performance effects found with anxiety are attributable to the increases in muscle tension that accompany anxiety. Not only do both muscle tensions and anxiety have

essentially the same effects upon measures of response stereotypy, but both of these variables interact in the same way with scores on the psychasthenia scale from the MMPI. Under control conditions, Ss with low psychasthenia scores approximate equal usage of each of the eleven available responses much more closely than do Ss with high scores on the psychasthenia scale. The low Ss also show fewer sequential dependencies in the temporal organization of their responses. However, when either anxiety or muscle tension is experimentally introduced, the low scoring Ss become more biased or stereotyped in their use and organization of responses while the high scoring Ss become less so.

Very similar interactions between anxiety and psychasthenia scores have previously been reported. Deese *et al* (3) found that the introduction of electric shocks on a nonsense syllable learning task resulted in a decrease in the number of correct responses by the low psychasthenia Ss and an increase by the high Ss. Eriksen (4) found that high and low psychasthenia groups differed on a measure of stimulus generalization when random electric shocks were administered during the task. While the nature of the responses measured in these two studies and in the present experiment differ in terms of such variables as the amount of competition between responses and the degree of response learning or differentiation, all three experiments are consistent in showing that whatever is measured by the psychasthenia scale is a significant determiner of individual differences in performance under anxiety.

The literature on the effects of muscle tension upon performance measures also contains evidence of marked individual differences in response to induced tension. An interaction effect quite similar to that obtained in the present study is reported by Bills and Stauffacher (1) and by Stauffacher (10). The Ss who are poor performers under control conditions on a learning task are reported more likely to improve in performance under induced tension than Ss who are initially good performers. They also report that some Ss show a complete disruption of performance when tension is introduced.

The general research on muscle tension has suggested that there is an optimum level of muscle tension for the performance of a given task with greater or lesser degrees of tension resulting in poorer performance (2). Stauffacher (10) has made use of this general finding

in proposing a theory to account for the interaction between slow and rapid learners and induced tension. Essentially, the theory holds that slow learners characteristically maintain a level of tension below the optimum while rapid learners are already operating at this optimum level. The effect of induced tension would then be to improve the performance of the slow Ss by moving their tension level up toward the optimum and to impair the performance of the rapid Ss by increasing their tension level beyond the optimum.

The results of the present experiment are consistent with Stauffacher's hypothesis if the psychasthenia scale can be assumed to measure individual differences in muscle tension. Thus, Stauffacher's rapid learners would correspond to the present high psychasthenia group and his slow learners to the low group. Such an assumption would appear justified in view of the general clinical impression that Ss with high psychasthenia scores are more tense than low scoring Ss. Also the high positive correlation between the psychasthenia scale and the Taylor scale of manifest anxiety (3, 4, 6) is consistent with such an interpretation, particularly since high scores on the anxiety scale have been found to characterize more rapid learners on simple tasks (11).

The relation of the present results to Stauffacher's findings is limited by the dissimilarity between the tasks used in the two studies, although there is a greater similarity between the task in the present experiment and the learning tasks employed by Stauffacher than is at first apparent. Previous research has shown that Ss do not use numbers in a purely random fashion (12). Not only are there number preferences but there are sequential relations in their response series. These response biases are in part at least unique to the individual Ss and may be considered to reflect differences in past learning in the relations between numbers and their use. Examined from this point of view, the present results suggest that under the control condition the high psychasthenia group shows a greater influence of this prior learning than does the low group; whereas, under the tension or anxiety conditions, the effects of this prior learning become more pronounced in the low psychasthenia group and less obvious in the high group. When stated in this manner, performance in

the present study is somewhat comparable to that reported by Stauffacher.

SUMMARY

The stereotypy and bias with which Ss used the number series 1-11 in a pseudo-ESP experiment was determined under three experimental conditions: (a) experimentally induced anxiety, (b) experimentally induced muscular tension, and (c) a control condition. Twenty-four Ss were selected with scores in the highest quartile on the psychasthenia scale from the MMPI and another twenty-four with scores in the lowest quartile. Eight Ss from each extreme group were randomly assigned to each of the three experimental conditions. The responses generated by the Ss on the pseudo-ESP task were analyzed for differences in frequency of usage of the different numbers and for sequential dependencies among the responses.

A significant interaction was found between the high and low psychasthenia groups and the three experimental conditions. Under the control condition, the high group was found to be more stereotyped and predictable in their use of the number responses, but under the anxiety and muscle tension conditions this relationship was reversed. Both muscular tension and anxiety had the same effect upon responses relative to the control condition and interacted in the same manner with psychasthenia scores.

It was concluded that at least some of the effects of anxiety upon responses and perform-

ance may be attributed to the increased muscular tension that accompanies the anxiety.

REFERENCES

1. BILLS, A. C., & STAUFFACHER, J. C. The influence of voluntarily induced tension on rational problem solving. *J. Psychol.*, 1937, **4**, 261-272.
2. COURTS, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347-367.
3. DEESE, J., LAZARUS, R. S., & KEENAN, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55-60.
4. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561-565.
5. ERIKSEN, C. W., & WECHSLER, H. Some effects of experimentally induced anxiety upon discrimination behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 458-463.
6. ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135-137.
7. GARNER, W. R., & MCGILL, W. J. Relation between uncertainty, variance, and correlation analysis. *Psychometrika*, 1956, **21**, 219-228.
8. MATARAZZO, J. D., ULETT, G. A., & SASLOW, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, 1955, **53**, 79-95.
9. MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204-220.
10. STAUFFACHER, J. C. The effect of induced muscular tension upon various phases of the learning process. *J. exp. Psychol.*, 1937, **21**, 26-46.
11. TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-64.
12. YULE, G. U. On reading a scale. *J. Roy. Statist. Soc.*, 1927, **90**, 570-587.

Received July 21, 1956.

LEADING AND FOLLOWING PSYCHOTHERAPEUTIC TECHNIQUES WITH HYPNOTICALLY INDUCED REPRESSION AND HOSTILITY¹

JESSE E. GORDON²

The Pennsylvania State University

PSYCHOTHERAPY research has in general been refractory to experimental methods. A set of procedures is needed by means of which laboratory research into the influence of a variety of patient, therapist, and outcome variables may be conducted. These procedures, however, must be such that results may be generalized to clinical practice. The ecological validity required of such research makes it necessary that subjects be permitted a wide range of responses, as are patients in therapy. Such requirements make experimental control and direct behavioral measurement difficult.

These problems are not so acute when the experimental variables are global ones such as directive vs. nondirective therapy, therapeutic success as defined by various criteria, and patient "maladjustment." However, systematic investigation aimed at building up a body of knowledge regarding therapeutic practices, instead of simply "proving" one school of thought to be superior to another, requires more refined control and manipulation of discrete, unitary variables rather than global ones. When a body of knowledge has been acquired in which the relationships among particular therapeutic practices and specific responses by patients are known, then a therapist may decide what effect he wishes to have on a particular patient and select his therapeutic techniques accordingly. Although this state of affairs may be an unattainable goal, it seems to be the only possible ideal at which the scientific therapist can aim.

With these considerations in mind, a set of experimental procedures was developed which seemed to possess ecological validity, which permitted a measure of control over the major client, therapist, and outcome variables, and which permitted the use of direct behavioral measures as well as psychometric methods.

¹ This article is based upon a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Pennsylvania State University. The writer wishes to acknowledge his indebtedness to Dr. George M. Guthrie and Dr. William U. Snyder.

² Now at the University of Wisconsin.

The experimental procedures involve the use of hypnosis to make normal subjects believe and act as if they had undergone a standard, mildly upsetting experience. The Ss have freedom to react to the experience in their own idiosyncratic ways. Reviews by Weitzenhoffer (9), Hull (5), and Rapaport (8) present much direct and indirect evidence to support the validity of generalizing from such hypnotically stimulated reactions to reactions actually encountered in clinical patients.

The use of a standard mildly upsetting experience, hypnotically induced, puts control over an important patient variable into the hands of the experimenter. This kind of control is practically and theoretically impossible in field research in psychotherapy. The acquisition of such control by the experimenter represents the major advantage of this type of research over therapy research conducted in the clinic with a patient population.

Instructions to the hypnotized Ss were designed to make them believe that they had been bumped into by a physical education teacher in high school and unjustly blamed for the incident. These instructions were patterned after those used successfully by Counts and Mensh (1). In addition, it was suggested to S that the therapist who was to interview him after he awakened bore some resemblance to the physical education teacher. Posthypnotic amnesia for the upsetting experience was also induced. These hypnotic instructions are based on a neurosis model which characterizes neurotically inappropriate behavior as resulting from an unconscious identification of the patient's present situation with an unpleasant and often unremembered experience in the patient's past.

The independent variables of this experiment were therapist's Leading (asking questions, making suggestions, etc.) and the therapist's Following (restating, reflecting, clarifying). These two kinds of therapist verbal behavior were selected for experimental comparison on the basis of investigations which indicate that they reflect a basic split among schools of psychotherapy (2, 4, 7). Roughly

paralleling this split regarding the use of Leading and Following behaviors is a split in what is regarded as the important outcomes of therapy. Those who use Leading techniques tend to regard the lifting of repressions as the aim of therapy. Those who use Following techniques generally disclaim interest in the historical approach to therapy and focus instead on improvement in the patient's present interpersonal functioning as the aim of therapy. It seems appropriate, therefore, to employ measures of both types of outcome variables, repression and interpersonal functioning. The hypnotic instructions to forget the "traumatic" experience produce behavior analogous to repression, making possible the measurement of the lifting of repression. The inclusion in the hypnotic instructions of a suggested similarity between the teacher in the traumatic experience and the therapist provides a situation analogous to transference, and thus makes possible some measurement of the inappropriate, "neurotic" interpersonal behavior of *S*.

The specific purpose of the present research, then, was to test a set of procedures for the experimental investigation of psychotherapy by applying them to a study of the effects of therapist Leading and therapist Following on hypnotically induced repression and on neurotic interpersonal functioning.

METHOD

Procedure

Sixty-two male college students volunteered to participate in "an experiment on hypnosis." Two group and then one individual training and screening sessions were held, out of which 26 of the volunteers emerged as susceptible enough to obey a hypnotic suggestion to be amnesic for trance events. Of this group, the data on eight subjects were lost as a result of faulty recordings of the interviews. Thus, 18 *Ss* remained in the two experimental conditions, Leading and Following.

The *Ss* were appointed for the individual experimental sessions following the third training and screening session. During this third session they were instructed, while hypnotized, to imagine themselves in a situation in high school in which a physical education teacher accidentally bumped into the *S* in a hallway, then blamed *S* for clumsiness and threatened reprisal. The *S* was instructed not to remember the incident after he was awakened and was told to have a daydream about it upon a signal. This served as a check on the acceptance of the hypnotic suggestions and as practice in covertly playing the role of an unjustly blamed person. Following this test, the graduate student "therapist" to whom *S* had been assigned was intro-

duced to the subject. The *S* was then rehypnotized and told that the incident with the physical education teacher had actually occurred. In addition, it was specified that the physical education teacher looked "something like Mr. ———, the man who will be interviewing you soon." The instructions used to induce belief that the incident had actually occurred were those used with success by Erickson (3). In addition to Erickson's instructions, it was further specified that though *S* would not be able to recall the experience when he awoke, it would bother him, and he could spend the interview hour trying to figure out what was bothering him. The instructions made it permissible to recall the experience during the course of the interview. Before being awakened, *Ss* were made amnesic for the entire session, including the introduction of the therapist.

The *S* was then awakened and brought into the interview room for a 50-minute interview, which was tape recorded. Following the interview, *S* was again rehypnotized, permitted to recall the content of the previous hypnosis, and disabused of the experience.

The therapists were nine advanced graduate students in clinical psychology at the Pennsylvania State University. Each therapist was in possession of a statement describing the acceptable therapist verbal behaviors in the Leading and Following conditions. These were explained and discussed more fully in a group meeting with the therapists. However, the therapists were not told the nature of the dependent variables or that a paramnesia had been induced in the *Ss*. Each therapist appeared twice, once in each condition, in an effort to control intertherapist differences. The order in which each therapist appeared in the conditions was counterbalanced.

Typescripts were prepared from the tape recordings of the interviews. Therapist statements on all even-numbered typescript pages were categorized by the *E* as Leading, Following, and Miscellaneous as a check upon the extent to which the experimental conditions were realized.

Repression. The content of the experimental "traumatic experience" had been broken down into meaning units, and criteria for scoring recall of the units were prepared. Three independent graduate student judges scored the typescripts, using these criteria to obtain scores for an Amount Recalled measure of the extent to which repression had been lifted. The means of the recall scores assigned by the judges constituted the scores on this variable. Agreement among the judges was high ($r = .94, .98, .93$).

The judges were also asked to indicate the statement in each typescript containing each *S*'s third scorable recall item. By locating these *S* recall statements on the tape recordings, it was possible to record the amount of elapsed time from the beginning of the interview, thus yielding a Recall Time score of the speed with which successful lifting of repression began to occur.

One other measure related to the lifting of repression was designated Recognition Time. To obtain this measure, the judges were asked to check the first *S* statement in each typescript which contained a verbalized awareness that the therapist, or someone like him, was familiar to *S*. Again, by locating the checked *S* statement on the tape recordings, it was possible to measure the elapsed time from the beginning of the

interview to the point at which this kind of insight began to develop.

Hostility. Hostility related behavior of the Ss was measured by means of rankings. Verbal descriptions of three aspects of hostility were given to three graduate students in a course in therapeutic interviewing at the University of Wisconsin, who then ranked the typescripts on each aspect. These hostility variables were (a) Extrapunitiveness, in which S attempted to convey his hostility to the interviewer in a direct manner, (b) Passive Hostility, in which hostility was expressed through negativism, uncooperativeness during the interview, and general resistance to the therapist, and (c) Disguised Hostility, in which S's hostility was communicated in symbolized or other indirect fashion, such as discussion of hostile dreams, overemphasis on aggressive feelings toward others, etc. Coefficients of concordance revealed significant agreement among the judges in their rankings on the first two variables but not on Disguised Hostility, which was therefore not included in further statistical analyses. Each S's rank on the two hostility variables was the mean of the ranks assigned to him by the three judges. Kendall's tau was computed between the rankings on Extrapunitiveness and on Passive Hostility and was found to be nonsignificant. The two indexes were therefore regarded as independent.

RESULTS

The results of the classification of the therapists' verbal behaviors are presented in Table 1 in terms of frequencies and proportions.

Repression

Amount Recalled. Significance testing was in terms of the differences between each therapist's Leading and Following interviews. Six differences showed greater Amount Recalled in the Leading condition with two differences favoring the Following condition and one of zero value. A randomization test of the differences, described by Moses (6), indicates that this distribution of differences just falls short of significance at the 5 per cent level with a two-tailed test.

Recall Time. Analysis of Recall Time was again in terms of the differences between each therapist's Leading and Following interviews. Three differences showed faster recall in the Leading condition; five differences favored the Following condition, and one difference was zero. A randomization test of the differences indicates that this distribution does not depart significantly from chance ($p > .05$).

Recognition Time. The differences in Recognition Time between each therapist's Leading and Following interviews indicated more rapid recognition in the Leading condition five times,

TABLE 1
MEAN FREQUENCY AND PROPORTION OF THERAPIST
LEADING AND FOLLOWING STATEMENTS IN THE
LEADING AND FOLLOWING CONDITIONS

Measure	Leading Condition		Following Condition	
	Leading	Following	Leading	Following
Frequency:				
Mean	26.88	10.22	10.55	27.22
Sigma	12.22	8.48	10.93	13.64
Proportion:				
Mean	.64	.21	.20	.65
Sigma	.21	.16	.14	.19

three differences favoring the Following condition, and one difference of zero. A randomization test of the differences indicates that this distribution does not depart significantly from chance ($p > .05$).

The correlation between Amount Recalled and Recall Time was $-.78$, between Amount Recalled and Recognition Time $-.55$, and $+.50$ between Recall Time and Recognition Time. All three correlations are significant at the .05 level. The very high correlation between Amount Recalled and Recall Time probably represents the extent to which these variables measure S's ability to repress and to conform to the hypnotic suggestions. On this assumption, an analysis of variance of errors of estimate in the regression of Amount Recalled on Recall Time was carried out. In effect, this procedure partialled out intersubject variability in conformity to the hypnotic suggestion. The results of this analysis are presented in Table 2³.

³It is important to recognize that this analysis is not an analysis of covariance. In this analysis, the matching variable (Recall Time) was obtained under the experimental conditions and was presumably affected by the conditions. In covariance analysis, the matching variable must be measured before the experimental conditions are imposed. The analysis of variance of errors of estimate as carried out here is, in effect, a test of the significance of the difference between the correlation between Amount Recalled and Recall Time in the Leading condition and in the Following condition. The analysis demonstrates that the correlation is very high in the Leading condition ($-.93$), thus the interpretation that Ss are recalling to the limits of their abilities in this condition; it is much lower in the Following condition ($-.63$), thus yielding larger errors of estimate and indicating that other factors have influenced the Amount Recalled, reducing it below the limits of the S's ability to recall as this ability is estimated from the Recall Time variable.

TABLE 2

ANALYSIS OF VARIANCE OF ERRORS OF ESTIMATE
IN THE REGRESSION OF AMOUNT RECALLED
ON RECALL TIME

Source	df	Mean Square	F
Between Treatment	1	62.94	12.87**
Among Therapists	8	17.66	3.61*
Residual	8	4.89	
Total	17		

* Significant at .05 level.

** Significant at .01 level.

Table 2 indicates that the errors of estimate in the regression of Amount Recalled on Recall Time were significantly larger in the Following condition when differences among therapists are eliminated from the comparison. It would appear that in the Leading condition, Ss tended to verbalize the "repressed" material more nearly to the limits of their ability to recall; whereas in the Following condition, other variables beyond the basic recall factors entered into the determination of how much recalled material was verbalized, thus producing larger errors of estimate and less verbalized recall. Differences among therapists were also found to be significant, which probably reflects differences in therapeutic skill.

The interviews were divided into those which were the therapists' first interviews and those which were second, and then analyzed in order to test the effect of therapist experience in this sense. Randomization tests of the differences indicate that first interviews produced significantly more recall than second interviews ($p < .05$), but Recall Time and Recognition Time variables were not significantly different from first to second interviews.

All interviews were then combined and ranked in terms of the number of Leading statements, the proportion of Leading statements, and on the three repression variables. Kendall's tau statistic was computed between all pairs of independent and dependent variables. The only relationship found that approximated significance was expressed by a tau coefficient of .31, with an associated probability of .075, between Amount Recalled and proportion of Leading statements.

Hostility

Mann-Whitney *U* tests indicate that there was no significant tendency for Leading or Following interview Ss to be ranked higher on

either hostility variable. Chi square was computed between classification of the interviews as above or below the median rank on Extrapunitiveness and their classification in the same terms on Proportion Leading. The value obtained after correction for continuity was 3.56, just short of significance at the .05 level. Similar analyses between Extrapunitiveness and Amount Leading, and between Passive Hostility and Amount and Proportion Leading, yielded definitely nonsignificant results.

The rank order correlation between rankings on Extrapunitiveness and on Recognition Time was .44, which is significant beyond the .05 level. Passive Hostility was not significantly correlated with Recognition Time, and neither of the two hostility variables was found to be significantly related to the other repression variables. The relationship between Extrapunitiveness and Recognition Time would be expected on the basis of the nature of the experimental instructions regarding the similarity between the therapist and the teacher. An *S* who quickly "recognized" his interviewer would be expected to be, and was, more hostile toward him than one whose recognition was delayed. One cannot decide, on the basis of these data, whether the recognition led to the hostile expression or *vice versa*.

As in the analysis of the repression variables, the typescripts were divided into those for the therapists' first interviews and those for their second, and analyzed for any relationships between sequence and hostility. A Mann-Whitney *U* test yielded a probability of less than .01, indicating that Ss were significantly more extrapunitive in the therapists' second interviews than in their first. It will be remembered that there was also significantly less recall in these interviews.

DISCUSSION

The results of this experiment indicate that Leading therapist behaviors tended in the samples used to be more efficient in obtaining as much verbalization of "repressed" material as *S* was capable of. A review of the typescripts lends support to this interpretation; in the Following interviews one notes several instances in which Ss failed to pursue what appeared to be a productive line of thought, ostensibly because of the lack of stimulation from therapist.

The finding that therapists' first interviews were more productive than second interviews,

in terms of the amount of recall, appears contrary to what one might expect. Second interviews should have reflected the insights gained by the therapists in their first interviews, and therefore might have been expected to be more productive. A suggested explanation for the findings is that the poorer results in the second interviews arose as a result of generally incorrect hypotheses developed by the therapists on the basis of their experiences in the first interviews. None of the therapists were able correctly to identify the nature of the experiment or the role of hypnosis after the first interview. Considering the many things about which Ss spoke in the interviews, it was highly probable that therapists would give undue weight to irrelevant verbalizations in forming hypotheses about the experiment and the Ss. This may be a function of the relative lack of experience of the graduate students who served as therapists. If this interpretation is valid, the danger of premature diagnostic hypotheses becomes apparent.

The significantly greater Extrapunitiveness of Ss in the therapists' second interviews is of interest. The relationship between Extrapunitiveness and Recognition Time indicates that the hostility is at least in part a product of the experimentally induced negative transference. The data do not permit a clear understanding of the relationship between this negative transference and the serial position of the interviews. It will be recalled that Ss in the therapists' second interviews had significantly lower Amount Recalled scores. Possibly the direct hostility in these interviews inhibited the recall, but it appears equally plausible that the unresolved repressions in the second interviews produced tensions in Ss expressed as Extrapunitiveness. Neither of these hypotheses was supported by the nonsignificant correlations between Amount Recalled and Extrapunitiveness. However, the experimental design used in the present research provides a method for testing these hypotheses more systematically in the future.

Many of these suggested conclusions stem from inferences drawn from relatively unreliable data. However, the experimental technique employed in the present study has demonstrated its usefulness in studying psychotherapy variables and generating new hypotheses. That these procedures bear striking similarities to clinical practice is demonstrated by the behaviors of the Ss, who appeared genuinely anx-

ious about the "repression." They seemed highly motivated yet natural in their associations, defenses, topics of conversation, etc.

Of incidental interest in reviewing the transcripts are the chains of associations by which Ss approached and eventually resolved the repression. These associations were natural, involving the telling of past experiences in the Ss' lives related to the "repressed" incident. For example, one S began the interview by describing his hostility and guilt when, as a small boy, his mother had punished him unjustifiably. She later overheard him expressing his anger toward her to a playmate. The relevance of this memory to the induced experience is obvious. It is interesting that this S eventually recalled most of the induced experience but could not correctly identify the relationship between the therapist and the physical education teacher. However, he did manufacture a plausible though incorrect reason for his feelings toward the therapist and experienced a real sense of relief when he hit upon this "reason." This observation suggests that true insight may not be a necessary prerequisite for relief from anxiety. It is quite possible that reasonable rather than accurate schematizations of experience contribute most to therapeutic effectiveness. Other interesting phenomena encountered in the interviews were dynamic determiners in word choice ("I'm sure we've *bumped into* each other somewhere or other") and the many ways in which hostility toward the therapists was expressed.

Other examples of realism in the interviews are provided by one S who spent the entire interview time "spontaneously" discussing difficulties in controlling frequent and intense anger. Another S experienced much embarrassment in relation to the therapist, whom he eventually discovered reminded him of his high school physical education teacher. He described the teacher as a father figure who disappointed him by not always behaving in the perfect way in which the S idealized him. The specific incident which S finally described as support for his disappointment was the induced one. Thus, there is considerable clinical evidence that the reactions of the Ss to the induced experience were idiosyncratic and clinically realistic.

The procedures used here appear to be practicable, and they potentially provide reliable and quantifiable information on a number of important psychotherapeutic issues: the rela-

tionships between the lifting of repressions and transference and hostility, the effects of therapist experience, the interaction of therapist and client personalities, the effects of therapist's sex on male and female Ss, etc. The procedures in their present form, however, are not without their limitations. The use of hypnotically susceptible Ss places an important restriction on the extent to which experimental results may be generalized. The restriction of this kind of study to a single interview with each S makes it difficult to study such procedures as the free association method, which requires fairly extensive training of S. Further, the traumatic neurosis model implied in the experimental design does not take into account the observation that neuroses occur only as a result of many such experiences in persons of particular personality structures.

Finally, there are problems in controlling therapist behavior. It is fundamental to the present study that the independent variable is a unitary, easily defined verbal response of the interviewer. An attempt was made to select for study a verbal response similar to those frequently made during psychotherapy. However, psychotherapy cannot be defined strictly by such limited verbal behaviors. The purpose of the present research was *not* to study psychotherapy; rather it was to provide a framework for studying variables involved in psychotherapy. This is an important distinction, and one which makes it necessary to select meaningful but well-defined variables while controlling as many associated contaminating variables as possible. As long as experimental research is concerned with different kinds of therapy, it will suffer from such problems as those involved in having therapists artificially change their professional styles, difficulties related to the experience of therapists, etc. Such problems may be largely avoided when research is conceived as chiefly concerned with particular classes of verbal response in face-to-face interactions.

These considerations emphasize the apparently essential limitations of experimental research in psychotherapy in that complexity and richness may be sacrificed in order to obtain reliable answers. However, there is still much to be profitably learned about the kinds of variables included within the relatively wide limitations of these experimental procedures.

Inasmuch as this method makes possible the reliable measurement of fairly discrete variables, it has value for current investigators. It may be important to find answers for the questions that can be confidently answered at present before going on to find less reliable answers to more difficult questions.

SUMMARY

A set of experimental procedures for use in psychotherapy research was tested. These procedures involved the use of hypnosis for the induction of miniature neuroses in normal individuals. Therapist Leading and Following, operationally defined, were studied for their effects on hostility and on the lifting of repressions in one-hour therapeutic interviews.

The findings indicate that Therapist Leading is more effective than Following for lifting repressions. It was also found that therapist experience which might lead to the formation of incorrect hypotheses about a client has a deleterious effect on the lifting of repression and is also associated with stronger expression of transference hostility by the client. The usefulness and limitations of the experimental procedures in providing possibilities for better control and measurement of psychotherapy variables were emphasized.

REFERENCES

1. COUNTS, R. M., & MENSCH, I. N. Personality characteristics in hypnotically-induced hostility. *J. clin. Psychol.*, 1950, **6**, 325-330.
2. DANKIN, D. G., & ROBINSON, F. P. Differences in "degree of lead" among experienced counselors. *J. counsel. Psychol.*, 1954, **1**, 78-83.
3. ERICKSON, M. H. The method employed to formulate a complex story for the induction of an experimental neurosis in a hypnotic subject. *J. gen. Psychol.*, 1944, **31**, 67-84.
4. GUMP, P. V. A statistical investigation of one psychoanalytic approach and a comparison of it with nondirective therapy. Unpublished master's dissertation, Ohio State Univ., 1944.
5. HULL, C. L. *Hypnosis and suggestibility*. New York: Appleton-Century, 1933.
6. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122-143.
7. PORTER, E. H. The development and evaluation of a measure of counseling interview procedures. II. The evaluation. *Educ. Psychol. Meas.*, 1943, **3**, 321-350.
8. RAPAPORT, D. *The organization and pathology of thought*. New York: Columbia Univ. Press, 1951.
9. WEITZENHOFER, A. *Hypnotism: An objective study in suggestibility*. New York: Wiley & Sons, 1953.

Received July 27, 1956.

THE EFFECTS OF PERSONAL AND SHARED THREATS UPON SOCIAL PREJUDICE^{1, 2}

SEYMOUR FESHBACH AND ROBERT SINGER

University of Pennsylvania

THE effectiveness of anxiety evoking communications in modifying attitudes has been explored in a number of investigations (3, 4, 8). Janis and Feshbach (3) were particularly concerned with demonstrating that communications which attempt to modify attitudes by means of threat may elicit defensive responses that reduce the intended effect of the communication. Clinical evidence (6, 9) suggests that, in addition to avoidant behavior, threatening situations frequently elicit hostility. The purpose of the present experiment is to determine the conditions under which threatening situations elicit hostility and affect the expression of social prejudice.

The immediate stimulus for this experiment was a study carried out primarily for pedagogical purposes during a seminar on group relations conducted by the senior author. The study was designed to test the simple hypothesis that the arousal of anxiety results in an increase in social prejudice. It was presumed that the anxiety would arouse hostility, manifested in turn in the expression of social prejudice. The similarity of this formulation to the application of the frustration-aggression hypothesis to social prejudice is apparent. Whereas in a study of social prejudice such as that of Miller and Bugelski (7), frustration was varied in order to produce variation in aggression, this demonstration involved manipulation of the degree of threat present in a stimulus situation. Schematically, the two respective formulations may be compared as follows: 1. Potentially frustrating situation: frustration—aggression—increase in social prejudice; 2. Potentially threatening situation: fear—aggression—increase in social prejudice.

The manipulation of threat was accomplished by varying the content of an interview conducted by the students in the seminar.

¹ This study was supported by a grant from the Albert M. Greenfield Center for Human Relations of the University of Pennsylvania. The authors especially wish to express their gratitude to Dr. Martin P. Chworsky, Director of the Center, who read the manuscript and made a number of helpful suggestions.

² A paper based upon this study was presented at the 1957 Eastern Psychological Association meetings.

Each student interviewed approximately six white adult respondents from the community such that half of his respondent sample were randomly assigned to the Control (neutral) interview and half to the Experimental (fear-arousing) interview. In the Control interview, the respondents were asked to give their opinions on the problem of the teacher shortage and overcrowding in the classroom. Their views were also requested on the issue of "pay television." They were then asked to complete a social prejudice questionnaire consisting of items selected from the California E Scale (1) which dealt primarily with attitudes towards Negroes; the questionnaire also included several general ethnocentric items from the E scale. A similar procedure was followed for the Experimental group with the important exception that questions about cancer and about the effects of an A-bomb attack and atomic radiation were used instead of the presumably neutral items asked in the Control interview. These questions were designed to arouse anxiety and constituted the experimental variable of the study. It was hypothesized that the Experimental subjects would respond in a significantly more prejudiced direction on the questionnaire than would the Control group.

Analysis of the experimental results revealed a significant difference between the Experimental and Control groups, but in a direction opposite to that predicted. Nineteen of the 27 subjects in the Experimental group obtained scores which fell below the median of the combined distribution while the scores of only 7 of the 25 Control subjects fell below the median. This difference in proportions yields a chi square of 9.3 which is significant at the .01 level of confidence.

This unanticipated result indicates the operation of a mechanism quite different from that originally postulated. Not only did the threat-arousal interview fail to produce an increase in social prejudice, but it resulted in an actual decrement in the expression of social prejudice. It is as if the threat interview evoked sympathy or a sentiment of "brothers under the skin" in its influence on the attitudes

of the white respondents toward Negroes. One could, of course, relinquish the threat-aggression hypothesis. However, an alternative possibility is proposed; namely, that the type of threat is the critical parameter. Specifically, it is postulated that a personal or ego threat results in an increase in hostility and in social prejudice while a "shared" threat results in a decrement in the expression of social prejudice.

The following criteria are offered for distinguishing between a personal and shared threat with respect to their possible influence upon social prejudice. A personal threat is one which is perceived by a subject as primarily directed toward himself rather than toward other individuals—particularly toward members of other ethnic and racial groups. Threats which elicit feelings of inadequacy and loss of status are likely to be personal threats. The possibilities of loss of job and bodily injury also fall within the category of personal threats if such threats are not shared by members of other social groups. Shared threats, on the other hand, are threats which are perceived as equally affecting members of both majority and minority groups. Shared threats do not discriminate between Negro and white, Jew or Christian. All groups share not only the danger but also the responsibility of coping with that danger. The prospect of war, hurricane, or flood are situations that are likely to have the characteristics of a shared threat. This is not to say that the prospect of war or hurricane must necessarily have the characteristics of a shared threat. Some individuals, particularly those with strong egocentric or paranoid tendencies, might interpret any impending disaster in very personal terms. Similarly, a threat that most individuals may perceive as personally directed may be reacted to by others in terms of a broader frame of reference. The examples of personal and shared threats cited above are intended as illustrations of the kind of situations that are likely to have personal or shared threat properties for most respondents in the general population.

Under the impact of a common threat, it seems reasonable to assume that one's reference group may become the population that is subjected to the danger. If this reference group now includes both Negro and white, whereas under ordinary stimulus conditions the reference group has been primarily the white population, then the social distance between white

and Negro should decrease, with a corresponding decrement in social prejudice. Social prejudice is at least, in part, a function of differences in group membership and such prejudice should be reduced to the extent that individuals perceive themselves as members of the same reference group; that is, as more alike due to the sharing of a significant common experience. One is reminded in this connection of the Gestalt notion of "common fate" as one of the factors determining perceptual organization.

METHOD

Materials

The type of threat was manipulated by varying the content of a mimeographed communication presented to the Ss. Each communication, whether Control, Personal Threat-Arousal or Shared Threat-Arousal, consisted of a brief discussion (approximately one-half page of single-spaced typing) of some issue. After reading this discussion, S answered a series of questions dealing with his attitudes and feelings regarding the issue.

The Control communication consisted of a discussion of pay television and of the length, frequency, and taste of television commercials. Three different Personal Threat-Arousal communications were used. One dealt with the frequency of divorce and the likelihood of making a satisfactory marital adjustment. The S was required to answer questions regarding the probability of his making a successful marriage and was asked to discuss any traits he had that might interfere with an adequate marital adjustment. A second Personal Threat dealt with the frequency of mental disorder and the respondent was asked to discuss his chances of becoming mentally ill. The third Personal Threat dealt with the dangers of fire and accident and the respondent indicated how he would react in the event of severe personal injury. Of the two Shared Threat-Arousal communications, one dealt with the danger of flood and hurricane while the second discussed the possibility of an atomic war.³ In both instances, the respondent was required to indicate how he would behave if actually exposed to such disasters. The questions asked of the Mental Health Threat group and of the Flood and Hurricane Threat group are presented in detail in order to illustrate the difference between the operations used to induce a personal threat and the operations used to induce a shared threat.

Mental Health Threat questions: 1. Have you ever thought of the possibility that you may develop some form of mental illness? (Discuss). 2. In case of mental illness, to what extent can you rely on your family for financial support and sympathetic understanding? (Discuss). 3. If you developed some form of mental disease and later recovered from it, do you feel that your occupational and social adjustment would be made more difficult due to once having been mentally ill? (Discuss briefly). 4. If you became psychotic would you prefer to be treated at home or at a mental hospital? (Discuss briefly).

³ Copies of the communications may be obtained on request from the authors.

Flood and Hurricane Threat questions: 1. About how many people do you think were killed as a result of floods and hurricanes in the past 12 months? 2. What would you do if the area you lived in were flooded or threatened by a hurricane? (Discuss). 3. What could you do if the people in your area, including close friends, couldn't get out of the danger area and all immediate help was cut off? 4. What do you think should be done for the people who would suffer great personal loss, physically, emotionally and financially, through an epidemic resulting from floods? 5. What do you think that people in your area can do to protect themselves against epidemics like the recent one in Lancaster?

The measure of social prejudice used in this study consisted of the six items from the California E scale dealing with attitudes toward Negroes, eight general ethnocentrism items, also taken from the E scale (1, p. 142) and 14 buffer items. Six alternatives ranging from "agree very strongly" to "disagree very strongly" were available for each item. The scoring of each item ranged from +3 for the most affirmative response to -3 for the most negative response. E-scale items are so constructed that agreement with an item is indicative of social prejudice.

Procedure

College students selected from introductory psychology classes served as subjects, and were assigned at random to the Control treatment, to one of the three Personal Threat treatments, or to one of the two Shared Threat treatment groups.

The social prejudice questionnaire was administered to the entire sample. Approximately four weeks later, the mimeographed communications were distributed to the Ss following a regular laboratory section meeting. Each laboratory section consisted of about 25 students. Because the experimental communications were silently read by each respondent, it was possible to distribute at random all communication forms to each laboratory section. Any single respondent was exposed, of course, to only one communication. After the Ss completed their answers to the communication, the social prejudice questionnaire was administered for the second time. Completed social prejudice questionnaires administered before and after the experimental treatments were obtained from 45 Control Ss, 37 Marriage Threat Ss, 24 Mental Health Threat Ss, 30 Fire and Accident Threat Ss, 25 Flood and Hurricane Threat Ss and 26 Atomic War Threat Ss.

It was predicted that each of the Personal Threat groups would show a significant *increase* in social prejudice in comparison to the Control group while each of the Shared Threat groups would show a significant *decrease* in social prejudice in comparison to the Control group.

RESULTS

Examination of the initial social prejudice scores indicated that the various treatment groups were homogeneous with respect to initial prejudice. Prejudice change scores were then calculated for each individual so that a positive shift reflects an increase in social prejudice, and a negative shift, a decrease. The

TABLE 1
COMPARISON OF MEAN IN PREJUDICE CHANGE SCORES
IN EACH THREAT GROUP VS. THE CONTROL GROUP

Treatment Group	M	σ	σ_M	t	p
Control (N = 45)	+0.33	5.09	.76	—	—
Marriage Threat (N = 37)	+3.30	7.4	1.2	2.11	<.05
Mental Health Threat (N = 24)	+3.04	6.1	1.2	1.92	<.06
Fire Threat (N = 30)	+3.03	6.2	1.1	2.10	<.05
Flood Threat (N = 25)	-1.64	5.8	1.1	1.51	>.10
Atomic War Threat (N = 26)	+2.60	7.3	1.5	1.38	>.10

mean and variance of the prejudice change scores for each of the experimental treatment groups are presented in Table 1. Except for the Atomic War Threat group, all changes are in the predicted direction. The difference between the Fire and Accident Threat group and the Control group is significant at the .05 level of confidence while the difference between the Mental Health Threat Group and the Control Group is significant at the .06 level. However, there is a significant difference in variance between the Marriage Threat Group and the Control Group which renders the *t* test inapplicable in this instance. The number of cases in each group falling above and below the median of the combined distribution was then determined. Twenty-four of the 37 Ss in the Marriage Threat group in comparison to 18 of the 45 Ss in the control group obtained scores falling above the median. This difference in proportions yields a chi square of 4.94, significant at the .05 level of confidence. The results for the Marriage Threat group then, in conjunction with the Mental Health Threat and Fire and Accident Threat data, lend strong support to the hypothesis that communications that arouse personal threats result in a significant increase in social prejudice.

In regard to the data bearing upon the effects of Shared Threats, the *t* test does not permit one to reject the null hypothesis of no difference between the Flood and Hurricane Threat group and the Control group. However, inspection of the data revealed a much greater proportion of decrements in prejudice in the Flood and Hurricane Threat group than in the Control group. The median of the combined distribution is 0, or no change, and 17 of the 25 Ss in the Flood and Hurricane Threat group in comparison to 18 of the 45 Ss in the Control group have prejudice change scores below the median, the difference yielding a chi square of 5.0, significant at the .02 level of confidence.

Thus the data for the Flood and Hurricane Threat group lend support to the hypothesis that the arousal of a Shared Threat results in a significant decrease in social prejudice. The data for the Atomic War Threat communication, however, fail to support the Shared Threat hypothesis. Because of a significant difference in variance between the Atomic War Threat group and the Control group, the number of cases falling above and below the median of the combined distribution was determined. Ten of the 26 Ss in the Atomic War Threat group and 26 of the 45 Ss in the Control group obtained scores falling below the median. While the difference is still in a direction opposite to that predicted, it is not statistically significant, chi square being 2.48.

DISCUSSION

The hypothesis that the arousal of a personal threat results in a significant increase in the expression of social prejudice is strongly supported by the data. The fact that each of three quite different communications yielded results consonant with the hypothesis provides additional confidence in the genuineness of the phenomenon. These results are readily accounted for by the theoretical analysis originally proposed; namely, the personally threatening condition arouses hostility which is then displaced on members of an out-group. While the data do not bear directly upon this issue, it is possible that the displacement of hostility in the form of increased prejudice has some functional significance in the sense of reducing the respondent's hostility or his insecurity. Thus the disparagement of the Negro implied by the increase in prejudice may partially compensate for the feelings of inadequacy aroused by the personally threatening communications.

There are undoubtedly a number of parameters related to the postulated threat-hostility-displacement sequence that warrant further study. The data indicate that hostility resulting from a personal threat is displaced on an out-group despite the lack of any conceivable connection between the out-group and the source of the threat. One may raise the question, however, as to whether it is necessary for some prejudice to be already present toward an out-group in order for such displacement to take place. Also, there are probably individual

differences in response to personal threats such that some subjects may tend to direct hostility toward themselves rather than toward other individuals or groups. It would be worthwhile to explore such personality variables as extrapunitiveness and intrapunitiveness in relation to their influence on the displacement of hostility. In addition, it may be useful to study the effects of different types of personal threats. Maslow's distinction between deprivation frustration and frustration involving psychological threat (5) suggests that ego threats may result in more hostility than deprivation threats although both types of situations fall within the category of personal threats as defined in this paper.

The findings for the Shared Threat-Arousal communications, while less conclusive than those with respect to the Personal Threat-Arousal Communications, are consistent with the hypothesis that exposure to shared threats results in a significant reduction in social prejudice. Of the two Shared Threat groups, the one exposed to the Flood and Hurricane Threat showed a reliable decrease in social prejudice while the group exposed to the Atomic War Threat did not show any significant change in comparison to the Control group. There are a number of possible explanations that may account for the failure of the group exposed to the Atomic War Threat condition to behave as predicted. First, the communication may not have been particularly credible or threatening for a college population. There is evidence, however, that it did have some effect; namely, it resulted in a significantly greater variance in the prejudice change scores in comparison to those of the Control group. Some subjects responded to the Atomic War Threat with relatively large increments in prejudice while others responded with relatively large decrements in prejudice. Thus the mean increment for those in the Atomic War Threat group who increased in prejudice is +6.4 and the mean decrement for those who decreased in prejudice is -6.3, while the corresponding figures for the Control group are +4.2 and -5.4.

While the atomic threat was effective in reducing prejudice in the original study, there are important differences between the threatening communications employed in the two studies. In the original study, the effects of the atomic threat were confounded with reactions

to the cancer threat.⁴ In the present study, the atomic threat was directly linked to a possible war with the Soviet Union. Because of this aspect of the communication, the responses to the general ethnocentrism items and the responses to the Negro items were separately analyzed. However, score changes in the Atomic War Threat group were similar on both sets of items. This correlation also held for the other threat groups with the exception of the Mental Health Threat Ss who manifested an increase in prejudice toward Negroes but did not show an increase on the general ethnocentrism items.

Another possible consequence of the linkage of the atomic threat to war with the Soviet Union is that some subjects may have responded with hostility to the Russians which they then generalized to other groups. Thus, there is experimental evidence (2) that, under certain conditions, the evocation of aggressive responses to one class of stimuli increases the probability of aggressive responses being made to other classes of stimuli. If this effect were operating in the present instance, one would have to modify the Shared Threat hypothesis to take into account possible differences between shared threats such as atomic war which may call for aggressive reactions, and shared threats such as flood and hurricane which call for more constructive group problem solving activity. Sherif (10) has provided evidence that the latter type of activity helps reduce aggressive tension between groups.

A question is then raised as to what were the essential characteristics of the Flood and Hurricane shared threat that brought about the reduction in prejudice. The initial theoretical discussion stressed the common impact of the shared threat, the fact that the majority and minority group members are equally threatened by impending disaster. However, the critical variable may be the recognition of the

⁴ While in terms of the present analysis the cancer threat would appear to be a personal threat, the impersonal manner of its arousal in the original study renders ambiguous its shared-personal threat status. The questions employed to elicit anxiety in regard to cancer were as follows: 1. What would you say is the most serious disease affecting American adults and for which an adequate cure has not yet been found? 2. What are the symptoms of cancer? How would a person know if he has cancer? 3. What are the chances of recovery from cancers such as lung cancer and stomach cancer?

need for shared, task-oriented behavior by both Negro and white in order to cope satisfactorily with the danger. It should be feasible to obtain evidence relevant to this issue by employing an experimental design in which the probable value of cooperative group activity with regard to reducing the threat is varied.

The results of the present study and those of the first experiment point up the importance of the interrelationship of motivational and cognitive factors in the determination of prejudiced behavior. They are consistent with the view that social prejudice is not a fixed trait or attitude that only awaits the proper conditions to reveal itself in its purest form. Rather, social prejudice is a consequence of contemporaneous cognitive and motivational factors and varies systematically as a function of variations in these determinants. The present data indicate that variations in the degree and type of threat are among the determinants of prejudiced behavior.

SUMMARY

The present experiment was designed to determine the effects of varying types of threatening communications upon the expression of social prejudice. Specifically, it was hypothesized that a personal threat results in an increase in prejudice while a shared threat results in a decrement in social prejudice. A personal threat was defined as one that is likely to be perceived by a subject as primarily directed toward himself while a shared threat is likely to be perceived as equally affecting members of majority and minority groups.

College students were assigned at random to either the Control treatment, to one of three Personal Threat-Arousal treatments or to one of two Shared Threat-Arousal treatments. A social prejudice questionnaire was administered to the entire sample. Four weeks later the experimental treatments were administered. The experimental variable was manipulated by varying the content of written communications presented just prior to readministration of the social prejudice questionnaires.

The main results were:

1. Each of the three Personal Threat groups showed a greater increase in social prejudice than did the Control group, the difference being significant at the .05 level for two of the groups.
2. Of the two Shared Threat groups, the

Flood and Hurricane Threat group showed a reliable decrease in prejudice in comparison to the Control group while the difference between the Atomic War Threat group and the Control group was not significant.

Since five of the six experimental predictions were confirmed by the data, the results were interpreted as being consistent with the theoretical distinction made between personal and shared threats. The possible influence of other variables on the response to personal and shared threats was discussed and the shared threat variable was further analyzed into a number of possible components.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. FESHBACH, S. The catharsis hypothesis and some consequences of interaction with aggressive and neutral play objects. *J. Pers.* 1956, **24**, 449-462.
3. JANIS, I. L., & FESHBACH, S., Effects of fear-arousing communications. *J. abnorm. soc. Psychol.*, 1953, **48**, 78-92.
4. JANIS, I. L., & FESHBACH, S. Personality differences associated with responsiveness to fear-arousing communications. *J. Pers.*, 1954, **23**, 154-166.
5. MASLOW, A. H. Deprivation, threat and frustration. *Psychol. Rev.*, 1941, **48**, 364-366.
6. MASLOW, A. H., & MITTELMAN, B. *Principles of abnormal psychology*. New York: Harper, 1951.
7. MILLER, N. E., & BUGELSKI, R. Minor studies of aggression: II. The influence of frustrations imposed by the in-group on attitudes expressed toward out-groups. *J. Psychol.*, 1948, **25**, 437-442.
8. MOLTZ, H., & THISTLETHWAITE, D. L. Attitude modification and anxiety reduction. *J. abnorm. soc. Psychol.*, 1955, **50**, 231-237.
9. SAUL, L. J. *The hostile mind*. New York: Random House, 1956.
10. SHERIF, M., HARVEY, O. J., WHITE, B. J., HOOD, W. R., & SHERIF, C. W. Experimental study of positive and negative inter-group attitudes between experimentally produced groups. Robbers Cave study. Norman: University of Oklahoma, 1954 (multilithed).

Received January 24, 1957. Prior publication.

CASE REPORT

THE WOLF

R. K. GREENBANK¹

Springfield, Delaware County, Pennsylvania

THE existence of fantasy playmates in young children is well known to child psychiatrists. Psychotics are also known to suffer from vivid hallucinations. This paper describes an adult nonpsychotic who continued the existence of a fantasied wolf from her childhood.

THE PATIENT

Miss L. was a twenty-four-year-old, single, white, Protestant, professional hospital worker who first sought psychiatric help for severe anxiety and phobic attacks. Face-to-face psychoanalytically oriented psychotherapy on a three-hours-a-week basis was used. Her Rorschach was interpreted as showing,

a fairly typical record of an hysteric. She shows a great deal of anxiety with a need to escape from her present environment. She is extremely immature emotionally, not realistic or logical. She feels herself appropriating unduly the time of others. Her personal relationships are extremely poor and her friends limited in number.

The patient's intellectual level was high average to superior.

"MY ANIMALS"

Early in treatment Miss L. spoke frequently of animals and with more affect than she typically used about humans. "In general, I like animals better than people. If people behaved like animals, this city wouldn't be such a madhouse. People are little, mean, and nasty; dogs aren't." "I feel like an animal on a leash." By her eleventh visit she could say of the therapist, "You watch me like I watch an animal. This is good because you understand me and are not excited by me."

She spent much time discussing pet dogs (real ones) she had owned in childhood. She was much involved emotionally in their deaths. She was particularly concerned with the clipping of one dog's tail. "I'm always close to my dogs. If I don't have one, I feel afraid."

By the 80th hour, she felt therapy was going deep and once said, "You better know what

you are doing, for if you take something away from me which I have, I'll die. I just won't be anymore."

After 100 hours of therapy the patient felt sufficient confidence to admit the existence of her fantasied wolf. She had kept this secret from her family, friends, and former therapist. She had previously feared the reaction of people if she discussed this aspect of her life. In a very guarded way over a period of several hours she began to reveal a particularly vivid and affectively colored fantasy life. She first mentioned her fantasy animals of childhood, then the present existence of "My horse and the wolf."

The patient had a large group of fantasied animals in childhood. These included a lion, a bear, a fox, squirrels, dogs, rabbits, etc., all of which were her friends. "They did not like people, but they liked me 'cause I'm different from other people." She lost all of these as she matured.

The two animals that stayed beyond her childhood were "the Wolf" and a horse. The horse can be mentioned briefly, since it never took an active part in the therapy. It pulled automobiles that she rode in; it ran beside trains on the way home. She often imagined riding on him. He was a big white horse which always wore a harness "because I like them." She did not bring him into the city because "horses should run free in the country."

DEVELOPMENT OF THE WOLF

As a child the patient played a game of "Big Black Mouse" with her younger brother. They placed all their possessions on a bed, then the fantasied big black mouse would appear, and they would run in terror to the bed where they were safe. "My brother did not enjoy this much because he did not know what was going on." There was much pleasure in this game but some tinge of fear.

One day while she was playing alone, the mouse got between her and the bed and drove her into a corner; it was going to bite her. The wolf then made its first appearance. It ran in

¹ Student, Philadelphia Psychoanalytic Institute.

the door and gobbled up the mouse and served as her protector thenceforth. She kept the wolf "because I like him and he liked me." "He's an old friend. I think he is as old as I am."

The wolf is "two and a half feet tall, not a pretty color, kinda grey, real nasty to people." He looks almost like a "laughing wolf" but "has great big teeth." "He doesn't wear a harness because he is wild." "He is big, real big. He used to be over my head. He stayed the same size but I grew." "He has lots of fur—too much for a wolf, really."

The wolf was partially under the patient's conscious control, i.e., she could make him come when she especially wanted him, but he often appeared without her asking for him. He often surprised her. "He feels the way I do. If I like a person, he does too; if I don't, he doesn't."

The wolf on at least two occasions was left with other people to care for them. Once she left him with her general physician (unknown to him) and once with her psychiatrist. She had noted the concern of her psychiatrist about a personal problem of his own. She made no comment at the time, but a week later when he felt better she casually asked, "Do you mind if I take my wolf with me? You don't seem to need him anymore." Further investigation revealed that she had felt that "You needed the wolf more than I did." Accordingly, she had left him to take care of the therapist.

In the latter part of her therapy, Miss L. had a real pet dog, Toby, which she had to leave in her apartment. This bothered her for she felt he would be lonely, so she sometimes left the wolf to watch the dog.

The wolf was a major source of protection to the patient. When it was necessary for her to have a major operation, she described the wolf following the stretcher as she went to the operating room. "He was snarling and vicious. He would have torn the doctors to pieces if they did anything wrong." She felt calm but said, "If it were not for him I'd have been scared to death."

The wolf had the same emotions as the patient. She often expressed her feeling through him. "My wolf is so sad since my boy friend went on vacation."

The wolf was a good indicator of the state of transference. At the first interview, the wolf sat firmly fixed between the patient and the therapist. He was "alert, watching every move

you made." Later he relaxed and often slept through the interview at the patient's feet. If the patient was under emotional strain or charged material was being approached, the wolf would raise his head and growl.

About a previous therapist, whom the patient had strongly disliked because of his aggressive manner, she said, "My wolf never rested when I was with Dr. M. He ran around the office, snarling, howling, and growling and his eyes were so red."

The wolf kept most frequently on the right side of the patient. He watched closely to see what the patient wanted. "He's glad to be there."

Early in therapy there was an occasion when the therapist found himself in close physical proximity to the patient on a train. He noted a sense of anxiety and concern in the patient and believed this was a result of the transference aspects of therapy and the physical closeness. He moved away and she showed signs of considerable relief. She never discussed this problem, although several openings were made by the therapist. After she had described the wolf to the therapist six months later, she then admitted that he had caused her a great sense of anguish, since he was standing on the wolf's tail. The fantasy animal was howling, but she felt unable to ask the doctor to move.

As the patient proceeded along well in her therapy, the wolf came less and less into notice. She fell in love with the president of a fraternity in a medical school in the same city. As this relationship developed, the wolf disappeared gradually from her life.

Earlier she had said, "It's nice to have you (the therapist) and the wolf, but I want a real person who talks to me, thinks about me, is nice to me, and whom I don't have to imagine." She found this in her boy friend.

After the wolf left the last time, she said, "He was chained up long enough to somebody who was unhappy and scared, and now he wants to run loose for awhile." But she missed him. "I do not know where my poor wolf is; I haven't had him for a long time. He must be around, but I do not know where."

DYNAMIC CONSIDERATIONS

Her fantasy animals, especially the wolf, probably prevented the development of a frank psychosis and allowed Miss L. to func-

tion on a neurotic level. This was, however, a fluid state. She developed paralysis, blindness, globus hystericus, tunnel vision, and *la belle indifférence* under moderate stress. If the strain became severe, especially with emotional deprivation, she regressed temporarily to more malignant symptoms such as withdrawal, paranoid thoughts, adult enuresis, and delusional thinking.

There was considerable emotional deprivation in her early life. Her neurotic mother and alcoholic father were unable to give her adequate love. Her loving and devoted grandfather died suddenly in the patient's presence. As a child, she was severely punished and rejected when she showed normal aggressive feelings.

By using the wolf as an auxiliary ego, she

was able to retain the essential aggression without suffering the guilt of being aggressive herself. Both the horse and the wolf, being symbols of masculine strength and power, provided her with a substitute for her weak father.

She, therefore, turned away from the emotional deprivation of her early life in the autistic pleasures to be gained in fantasy with her animals.

SUMMARY

The case of a young woman has been presented who continued into adult life a hallucinated wolf from her childhood. It remained her support and companion until she was able, through therapy, to develop a heterosexual relationship with an accepting male.

Received July 17, 1956.

CRITIQUE AND NOTES

A GUTTMAN SCALE FOR SEXUAL EXPERIENCE—A METHODOLOGICAL NOTE¹

LAWRENCE PODELL

The City College of New York

AND

JOHN C. PERKINS

*Veterans Administration Mental Hygiene
Clinic, Buffalo, N. Y.*

IN some investigations of sexual behavior, there is an implication that heterosexual experience cumulates in a certain sequential pattern. For example, Bromley and Britten (2, p. 290) employed the following item: "If you care for hot petting, how far do you go?" Kinsey, Pomeroy, and Martin (3) state that:

Petting techniques usually expand in a more or less standard sequence, as the partners become better acquainted. Beginning with general body contact, lip kissing, and the deep kiss, it advances to a deliberate manipulation of the female breast, to mouth contacts with the female breast, to manual stimulation of the male genitalia, and finally to oral stimulation of the female genitalia (3, p. 540).

This statement implies a cumulative sequence wherein a person who has participated in a "more advanced" behavior may be assumed to have experienced the "less advanced" behaviors. In other words, the statement suggests that a male who has had oral contact with the female breast has also engaged in lip kissing, deep (tongue contact) kissing, and manual stimulation of the female breast.

Does heterosexual experience in fact cumulate along a unidimensional continuum? If so, what is the sequence of behaviors? The present study is directed to these questions. To the knowledge of the authors, such a unidimensional cumulative scale of sexual behaviors has not been established in any studies thus far published.

METHOD

The responses used in this research derived from personal interviews. One hundred unmarried male college students, 50 with extremely unfavorable marital prediction scores and 50 with extremely favorable marital prediction scores on the Adams' Marital Happiness Prediction Inventory (1), volunteered to participate. The subjects were between 18 and 22 years of age. Fifty-eight reported themselves as Protestants, 24 as Roman Catholics, 15 as Jewish, and the remainder claimed no religious affiliation. Fifty-eight were raised in a rural environment, 35 in an urban environment, and the remainder in mixed settings.

RESULTS

The Cornell technique of Guttman Scale Analysis (4) was used to analyse the Ss' responses to

16 items concerning heterosexual behavior.² Of these items 15 were found to form a Guttman scale.³ The derived order of items appears in Table 1, with the number of errors and nonerrors presented for each item.

To determine the degree to which the derived scale deviated from a theoretically perfect Guttman scale, the coefficient of reproducibility was computed.⁴ Technically, the proper procedure in Guttman scaling calls for the elimination of items involving frequencies of 10 per cent or less and 90 per cent or more. Four of the 15 items (items 1, 2, 3, and 15 in Table 1) were in these categories. Although these items were used in Table 1 for illustrative purposes, they were not included in computing the coefficient of reproducibility for the scale. The coefficient of reproducibility is .935. Thus, knowing the subjects' scores, we can state those behaviors which the subjects have reportedly experienced with over 93 per cent accuracy. (The coefficient of reproducibility for all 15 items is .948.) The distribution of responses forms the

² The sixteen items were selected from a total of 271 items covering a wide variety of sexual attitudes and behavior. Questions 88 to 93 in the interview concerned (in the order given) close embrace, lip kiss, tongue kiss, fondling the clad breast, fondling the nude breast, kissing the breast, sucking the breast, caressing the female sex organs, fondling the male nipples, caressing each other's sex organs, fondling of the male sex organ to orgasm, bare genital contact without intromission, and interfemoral relations. Two items followed concerning the effects of petting. Question 96 concerned sexual intercourse. Fifty-eight items followed, after which the questions concerning oral contact with the male genitalia and female genitalia were asked. The actual wording of the questions may be found in the original study (5, pp. 253-266).

³ It was found that item 16, "In a necking or petting situation, has a girl ever played with your nipples" could be substituted for item 15, "oral contact with the female genitalia," and still fulfill the requirements of a Guttman scale. However, if both items are included in the attempt to derive a scale, either one, depending upon the ordering of respondents, will incur more errors than nonerrors and must be eliminated. In other words, either behavior is part of a unidimensional sequence, but not both. It is possible that different dimensions are introduced at this extreme point, but there were insufficient data to test this possibility.

⁴ The coefficient of reproducibility =

$$1 - \frac{\text{number of errors}}{\text{number of items} \times \text{number of respondents}}$$

¹ The data used in this paper were obtained in the course of a doctoral research by J. C. Perkins (5).

TABLE 1

A GUTTMAN SCALE OF PREMARITAL HETEROSEXUAL EXPERIENCE FOR 100 COLLEGE MALES

Se- quence	Type of Behavior Reported	Errors	Non- errors
15	Oral contacts with female genitalia	6	10
14	Oral contacts with the male genitalia	7	21
13	Interfemoral relations	7	34
12	Bare genital contact, without intromission	6	37
11	Sucking of the female breast	15	39
10	Manual manipulation of the male genitalia to orgasm, by the female	10	51
9	Sexual intercourse	8	49
8	Manual manipulation of each other's genitalia	3	63
7	Manual manipulation of the female genitalia, by the male	2	68
6	Kissing of the female breast	9	63
5	Manual manipulation of the nude female breast, by the male	1	73
4	Kissing with tongue contact	3	84
3	Manual manipulation of the clad female breast, by the male	1	91
2	Lip kiss	0	99
1	Embrace	0	99

parallelogram pattern characteristic of the Guttman scale.

DISCUSSION

In published researches, there seems to be a tendency to consider sexual intercourse as the most extreme behavior in a sequence of sexual behaviors. For example, the phrases "complete sex experience" and "go the limit" are apparently used as synonymous with sexual intercourse (7, pp. 61 and 247). The arbitrary judgment of any particular behavior as "most extreme" by an investigator may indicate personal biases on his or her part. This research, by objectively and empirically ordering premarital heterosexual experience along a unidimensional scale, would seem to suggest that sexual intercourse is not the most extreme type of sexual behavior, though it tends to be more extreme than some other types.

Not only do existing studies imply a cumulative sequence of sexual experience, but they also imply that persons experience the "less extreme" behaviors before the "more extreme" ones. This is an assumption which we cannot make; although a temporal pattern of experience will be cumulative, the existence of a Guttman scale does not prove a temporal sequence, e.g., the respondents who experienced item 12 may have done so before or after item 9. To demonstrate a temporal sequence, the questions would have to be phrased accordingly.

A meaningful interpretation of the sequence of

behaviors appearing in Table 1 awaits further research. That a single dimension of sexual experience was discovered rather than a plural number of dimensions (e.g., an oral dimension, a genital dimension) remains to be explained. It is possible that the ordering of the reported behavior is influenced by the factor of the "social acceptability" of the behavior. For example, the discrepancy in rank order between items 6 and 11 may be a matter of the respondent's perception of sucking the breast as an "infantile" behavior. (Another possible explanation for this rather outstanding discrepancy may be offered: The question concerning item 6, "kissing the female breast," did not specify the nipple of the breast and it was not ascertained whether or not the respondents interpreted the question to include the nipple. The same holds true for item 11, "sucking the female breast." However, it seems more probable that nipple contact was greater in the behavior of item 11 than in that of item 6; the difference in rank order of the items in the scale.)

It should be noted that Ss in this study were a highly selected sample. Therefore, the scale presented in this note may not be applicable to single men in general, nor even to single men whose backgrounds are similar to the backgrounds of the respondents used in this study. Nevertheless, the fact that a scale could be obtained from two extreme groups seems to suggest that it is feasible to develop satisfactory scales of sexual behavior, which may or may not be similar to the one presented here. Comparison of the incidences of sexual behaviors of this highly selected sample with the findings of Kinsey, Martin and Pomeroy (3) for a sample of corresponding age and educational level revealed little difference. The comparison may be found in Perkins' dissertation (5, pp. 267-268). Using other populations, we might expect to find different scale patterns or even the absence of unidimensionality. In the light of the studies by Kinsey et al (3) it would appear that in using populations which differed from the sample in socioeconomic status, educational level, or age range, we might expect differences in sexual experience (e.g., a differential distribution of petting behaviors by educational level) and the scalability of such experience.

In addition, it must be noted that the Guttman scale illustrated in this research need not be identical with a scale derived from the responses of females, even respondents with similar backgrounds. In fact, there is evidence available to suggest that this is not the case (6).

SUMMARY

The reported premarital heterosexual sexual experience of 100 highly selected college men was subjected to the Cornell Technique of Guttman

Scale Analysis, with a satisfactory scale forthcoming. It therefore appears feasible to order certain aspects of sexual behavior along a unidimensional cumulative scale, thus aiding in the definition of premarital sexual experience.

REFERENCES

1. ADAMS, C. R. *The marital happiness prediction inventory*. State College: Pa. State College, 1950.
2. BROMLEY, D. D., & BRITTEN, F. *Youth and sex*. New York: Harper, 1938.
3. KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male*. Phila.: W. B. Saunders, 1948.

4. GUTTMAN, L. The Cornell technique of scale and intensity analysis. *Educ. psychol. Measmt*, 1947, 7.
5. PERKINS, J. C. Marital prediction scores of college men relative to behavior and attitudes. Unpublished doctor's dissertation, Pennsylvania State Univer., 1954.
6. REEVY, W. R. The differential responses of women with a favorable marital happiness prediction and women with an unfavorable marital happiness prediction. Unpublished doctor's dissertation, Pennsylvania State Univer., 1954.
7. ROCKWOOD, L. T., & FORD, M. E. N. *Youth, marriage and parenthood*. New York: J. Wiley, 1945.

Received June 18, 1956.

LEVELS OF PERCEPTUAL ORGANIZATION AND COGNITION:
CONFLICTING EVIDENCE¹

E. W. PICKREL

Personnel Laboratory, Air Force Personnel and Training Research Center

RATHER low relations have generally been found between performance on perceptual and cognitive tasks (4, 5). However, Krech and Calvin (3) have recently reported a surprisingly high relation, a biserial correlation of .91 with a standard error of .07, between a culture-free perceptual variable and the culture-laden intelligence variable. Their perceptual variable was the extent to which a dotted pattern, presented tachistoscopically at .06 second, had been articulated; that is to say, whether a subject organized the stimulus (see Fig. 1) into a homogeneous pattern of equidistant dots or whether he left at least twice as much space between the rows as within the rows (or columns) of dots. As a measure of intelligence they selected the vocabulary test from the Wechsler Bellevue Intelligence Scale. It was suggested that this high relation was a demonstration of continuity between perceptual and cognitive processes, but the magnitude of their obtained correlation is at variance with that commonly found between such tasks. The purpose of this study is to reproduce their experiment to verify the results and observe the relation of this new measure to those perceptual variables most commonly used in studies of individual differences.

METHOD

Apparatus and Measures

The apparatus included a tachistoscope, stimulus cards, a "reproduction board" and a supply of small

black plastic beads. The reproduction board, plastic beads, and their use have been previously (3) described. The stimulus cards were presented by means of an electronically controlled Dodge-type mirror tachistoscope. The stimulus fields were illuminated by fluorescent lamps to a level of .61 foot candle and the illumination of the pre- and post-exposure field was matched with exposure field at the point of the observer's eye. The stimulus cards (see Fig. 1) were photographic copies from an original used by Dr. Calvin. The sole criterion of a successful reproduction² was that for any single trial there should be at least twice as much space

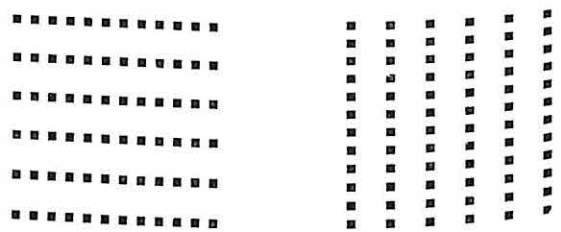


FIG. 1. STIMULUS CARD

between the rows as within the rows (or columns) of the patterns.

Following the tachistoscopic measure subject (S) was administered three of the Keystone Visual Skills tests, the vocabulary test taken from Wechsler (7), and three paper-and-pencil tests. The Keystone tests were for determining whether there was a significant relation between the perceptual phenomenon and visual skills, the paper-and-pencil tests for determining if articula-

¹ This report is based on work done under ARDC Project No. 7719, Task No. 17011, in support of the Research and Development Program of the Air Force Personnel and Training Research Center, Lackland Air

Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.
² As stated in a personal communication by Dr. Calvin.

tion was related to the three commonly used factor measures of individual differences in perceptual abilities. The tests selected from the Keystone battery are "near point" measures for lateral imbalance, fusion, and usable vision. The three paper-and-pencil tests are measures of the principal spatial ability factors that have been isolated (1). Pattern Comprehension Test BP803C yields a measurement of how quickly *S* can understand how the pattern of an object will fold into a solid figure. Rotated Figures Test BP112C is a measure of how quickly *S* can identify a stimulus card when it is rotated in two-dimensional space. Fish Bowl Test UT-19FB is a measure of the ability to determine rapidly spatial relations relative to the body orientation of the observer. The *S* is given a head-on view of an object floating in a fish bowl, then given a different view of the same object and must determine the new position to which he must project himself to obtain this second view of the object.

Subjects

To approximate the ability level of the sample used by Krech and Calvin, airmen with percentile scores of 85 and above on the Armed Forces Qualifying Test, a general intelligence measure, and AFROTC graduates attending the Officers Military School at Lackland Air Force Base were selected as *Ss*. One hundred and twenty-four *Ss* were used for the experiment.

Procedure

In a pilot study which used the directions and exposure time of the original experiment, no significant relation was found between performance on the perceptual and verbal measures. Similar results have been obtained concurrently by De Soto and Leibowitz (2). It was found from protocols that *Ss* had been seeking extraneous cues on which to focus their attention, and were not considering a solution as inherent in the stimuli being observed tachistoscopically. For this reason the instructions were rewritten to focus attention on the stimulus cards and accurate reproduction of perceptions.

Each *S* was given the following instructions at the beginning of his first experimental session:

I am going to present you with a test of visual acuity, that is how well you can observe. You must realize that visual acuity is becoming more and more important as our aircraft increase in speed. We are asking that you serve as guinea pigs, as pilots, for us while we are developing this measure. You will be presented with two symbols for a very brief exposure on the screen. For any given trial these symbols may seem to be alike or they may appear to be different. They may seem to change from trial to trial, or they may appear to be the same. After each presentation I want you to reproduce both of the symbols as accurately as you can by simply drawing them on this board with these beads as this would be a drawing of the five on dice (demonstrate). Now look into the box. You see a cross which is roughly in the center of a white field. One of the symbols will be on the left-hand side of the cross, the other on the right. They will

TABLE 1
VISUAL SKILLS TEST PERFORMANCE FOR SUBJECTS IN EACH GROUP

Levels of Perceptual Organization	Fusion			Convergence			Visual Acuity	
	N	Normal	Deviate	Over	Normal	Under	20/40 to 20/22	20/20 to 20/15
Nonarticulated Group	74	73	1	11	60	3	8	66
Articulated Group	50	49	1	7	39	4	6	44

TABLE 2
INTERRELATION OF THE PERCEPTUAL AND VERBAL MEASURES

	Paper-and-Pencil Measures			Level of Perceptual Organization
	Pattern Comprehension	Rotated Figures	Fish Bowl	
Vocabulary Test	.14	.03	.22*	.34**
Level of Perceptual Organization	.15	.10	.12	

* Significant at the .05 level.

** Significant at the .01 level.

Note.—Vocabulary—paper-and-pencil coefficients are product-moment correlations, the other four coefficients are biserial correlations.

be somewhat smaller than you can draw on the board, but make them proportional.

To repeat the directions, you will be presented with two symbols for a very brief exposure on the screen. The time is so short that I want you to say "ready" each time. Then the symbols will appear simultaneously. Remember that for any given trial these symbols may seem to be alike or they may appear to be different. They may seem to change from trial to trial, or may appear to be the same. After each observation draw both symbols as accurately as you can on the board and I will tell you whether your drawing is correct. Are there any questions?

Questions were answered in the same manner as in the original experiment. After each trial *S* was told if his percept was correct or incorrect for the purpose of directing closer and closer attention to stimulus detail. The procedure was continued until 16 trials had been completed.

RESULTS

It is evident from Table 1 that there was no significant relation between the control measures of visual skills and levels of perceptual organization. The measure of visual acuity yielded an r_{bis} of .05 with the articulated-nonarticulated pattern groups.

The interrelations of perceptual and verbal measures are presented in Table 2. The correlation between vocabulary scores and dichotomized levels

of perceptual organization was .34 with a σ_{bis} of .10. Using a test presented by Walker and Lev (6) it was found that the difference in magnitude between the .34 coefficient reported here and the .91 coefficient obtained by Krech and Calvin was significant at the .01 level.

Mean score on the vocabulary test for the 50 Ss who reached an articulated level was 28.64 with an SD of 3.76. Twenty of these 50 Ss had scores of 30.5 or higher on the vocabulary test, a cutting point which differentiated the articulated and non-articulated groups in the original experiment. Mean score for the 74 Ss holding to the homogeneous percept was 26.16 with an SD of 4.59. Fifteen of these Ss had scores of 30.5 or higher on the vocabulary test.

SUMMARY

Though rather low relations have generally been found between performance on perceptual and cognitive tasks, Krech & Calvin recently reported a surprisingly high relation between perception of a dotted pattern viewed tachistoscopically and performance on the Wechsler vocabulary test. The purpose of this study was to reproduce their experiment to verify the results and observe the interrelation of the measures employed to those perceptual variables most commonly used in studies of individual differences. A slightly negative relation was found between performance on the perceptual and cognitive tasks for an initial sample. Since an artificial problem contained within the original administrative directions seemed to distract Ss from the perceptual task, administrative directions were rewritten to focus each S's attention onto the perceptual problem. The Krech measure yielded a higher relation than the other perceptual measures

to verbal test performance and had a very low relation to the other perceptual tests.

The size of the correlation reported here between the Krech measure and verbal test performance was significantly lower than that obtained by Krech & Calvin in an earlier study. However, the correlations obtained in this study between vocabulary scores and perceptual test performance are of a size often found between perceptual and verbal measures in studies of individual differences and are more consistent with such information than that found by Krech and Calvin.

REFERENCES

1. ANDERSON, G. V., FRUCHTER, B., MANUEL, H. T., & WORCHEL, P. Survey of research on spatial factors. *Research Bull. AFPTRC-TR-54-84*, 1954, AF Personnel & Train. Res. Cent., Lackland Air Force Base, San Antonio, Texas.
2. DE SOTO, C., & LEIBOWITZ, H. Perceptual organization and intelligence: A further study. *J. abnorm. soc. Psychol.*, 1956, 53, 334-337.
3. KRECH, D., & CALVIN, A. Levels of perceptual organization and cognition. *J. abnorm. soc. Psychol.*, 1953, 48, 394-400.
4. PICKREL, E. W. Prediction of the trainability of slow learners from tests with a symbolic and non-symbolic content. *Res. Bull. AFPTRC-TR-54-82*, 1954, AF Personnel & Train. Res. Cent., Lackland Air Force Base, San Antonio, Texas.
5. THURSTONE, L. L., & THURSTONE, THELMA G. Factorial studies of intelligence. *Psychometr. Monogr.*, 1941, No. 2.
6. WALKER, HELEN, & LEV, J. *Statistical Inference*. New York: Henry Holt & Co., 1953.
7. WECHSLER, D. *The measurement of adult intelligence*. (2nd ed.) Baltimore: Williams and Wilkins, 1941.

Received June 25, 1956.

AUTHORITARIANISM OR ACQUIESCENCE IN BASS'S DATA¹

SAMUEL MESSICK

*Educational Testing Service and
Princeton University*

AND

DOUGLAS N. JACKSON²

Pennsylvania State University

SOME errors have been located in a recent article in this journal (1) which necessitate a reinterpretation of the results. In his study of authori-

¹ Most of these remarks were made in private correspondence with Bernard M. Bass. They are presented as a separate article at the suggestion of the editor in the interest of clear presentation of the corrections and interpretations.

² The authors wish to thank Dr. Milton Rokeach for first pointing out the disparity between Bass's data and his conclusions and Dr. Wayne H. Holtzman for participating in our joint critical analysis of Bass's article.

tarianism and acquiescence, Bass obtained a correlation of $-.20$ between the California F scale and an F scale rewritten so that the content of each item reflected a viewpoint opposed to the original (called the G scale). Bass used this correlation in a factor analysis of the F and G scales, placing reliabilities in the diagonals. The starting point for the analysis was a correlation matrix between F and $-G$ [Table 4 in (1)]. In this matrix the G scale had been reflected and $r_{FG(-)} = .20$ was used instead of $r_{FG} = -.20$. However, the notation for the factor matrix of Table 5 (1) was in terms of F and G, not F and $-G$. Since the

factor loadings in Table 5 reproduce a correlation of $r_{FG} = +.20$, the factor matrix labels should read F and -G just like the correlation matrix labels. Thus, F and -G, not F and G, have positive loadings on the first factor, necessitating a reversal in interpretation.

This outcome may also be seen by factoring the following unreflected correlation matrix with reliabilities in the diagonals:

	F	G
F	(.81)	-.20
G	-.20	(.50)

The factor matrix which reproduces these correlations is

	I	II
	Authoritarianism	Acquiescence
F	.77	.47
G	-.53	.47

and not Bass's Table 5, which is identical except for reversing signs in the second row and reversing the labels of the factors. Now, approximately three-fourths of the reliable variance of the F scale is apparently associated with an *authoritarianism* factor while only one-fourth is attributed to *acquiescence*.

Since these factors are based upon a single correlation coefficient, they are certainly not determined in the usual factor analytic sense. Ordinarily, at least three tests and preferably more are necessary to define a single common factor in Thurstone's system and at least five tests are required to determine two common factors (5). Also, since Bass used reliabilities instead of communalities in the diagonals, these factors involve both specific and common variance, thus making the interpretation particularly difficult. The label "acquiescence" attached to the second factor is a very reasonable interpretation, but there is no way of ascertaining whether Factor II might not be a second content dimension due to an inadequate reversal of the scales or perhaps even some confounding of specific content factors and acquiescence.

On the other hand, it should be pointed out that Bass was probably not attempting to isolate and establish psychologically meaningful factors; he was using factor analysis to *describe* the reliable variance of a scale and to show that one factor was not an adequate description. Another way of demonstrating this would be to use the reliabilities to correct for attenuation. If the unattenuated correlation between the two scales is significantly less than -1.00 (as it is in this case), then more than one factor is represented. Again, there is no way of determining the nature of the second factor from this single correlation coefficient. Although acquiescence provides the most likely hypothesis, the possible operation of additional content factors cannot be discounted.

Bass's interpretation of the analysis of variance presented in his article suggests that the variance attributable to acquiescence may still be significant. However, this analysis is based upon the assumption that if no response set existed, each individual's F scale score when combined with his G scale score should sum to zero. This would be true if the G scale items were perfect reversals in direction and equally potent in opposition to the corresponding F scale items. However, the various levels of opposition reflect imperfect reversals, and the different variances for the F and G scales indicate that they were not opposing parallel forms of the kind necessary to cancel scores. Since the basic assumption was not met, the analysis of variance is difficult to interpret.

There is also the possibility that agreements to corresponding original and reversed F scale items might indicate that the statements are not really opposite in meaning. The correlations between F and G range from -.52 to +.05 going from the best level of opposition to the worst. The r_{FG} of -.52 is in the direction predicted by consistent responses to content in the two scales, and it approaches the theoretical maximum correlation that would be expected given the present reliabilities. As the judged adequacy of the reversals decreases, r_{FG} becomes more and more positive, again as would be predicted in terms of content responses to items whose meanings are becoming less perfectly opposed. However, it should also be pointed out that the high negative correlation is associated with items reversed by inserting "not" or "no" into the original statement and that opposing responses would be expected to such item pairs. Acquiescence, on the other hand, has a much better chance of appearing in those items which have similar styles of expression yet are obviously opposed in meaning without being simple reversals. Such items might be clearly opposite semantically and still be judged less opposite than a simple reversal.

The point to be emphasized is that the over-all correlation of -.20 in this sample is in the content direction, but it is not as high as a complete content interpretation would require. Bass attempted to demonstrate using analysis of variance and factor analysis that more than a simple authoritarian-content dimension was necessary to account for this intercorrelation in light of the obtained reliabilities. An error in factoring resulted in too much variance being attributed to "acquiescence," but this error does not eliminate acquiescent response set as a possible determinant of F scale scores.

Even though Bass's conclusions do not necessarily follow from his data, his emphasis upon acquiescent response set in the F scale is important and is supported by other evidence. In some recent

studies, correlations between agreement to original F scale items and to various reversed forms were not found to be high and negative (4) as would be expected from consistent responses to item content and were sometimes found to be significantly positive (2, 3). Thus the operation of a consistent response set in the California F Scale has been repeatedly demonstrated and perhaps, as the latter two studies suggest, even a major portion of the variance may be legitimately attributed to acquiescence.

REFERENCES

1. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.

2. JACKSON, D. N. & MESSICK, S. J. A note on "ethnocentrism" and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132-134.
3. JACKSON, D. N., MESSICK, S. J., & SOLLEY, C. M. How "rigid" is the "authoritarian?" *J. abnorm. soc. Psychol.*, 1957, **54**, 137-140.
4. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215-221.
5. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. Chicago Press, 1947.

Received December 4, 1956. Prior publication

REPLY TO MESSICK AND JACKSON'S COMMENTS ON AUTHORITARIANISM OR ACQUIESCENCE

BERNARD M. BASS

Louisiana State University

IN THE light of Messick and Jackson's (3) corrections, for which I am grateful, and subsequent independent studies on the same subject, it appears that the hypothesis advanced in "Authoritarianism or Acquiescence" (1) was more satisfactory than the analysis used to test the hypothesis.

FURTHER COMMENTS ON METHOD

I should like to add the following methodological notes: The statement that any individual's F and G score cancel each other was not correct, but the analysis of variance was another way of showing that F correlated $-.20$ with G. At the same time, the analysis permitted the examination of the more complex effects of "levels of opposition."

The variance of the F + G scores is given by:

$$\sigma^2_{F+G} = \sigma^2_F + \sigma^2_G - 2r_{FG}\sigma_F\sigma_G$$

The relation or ratio between the F + G variance and the "within-scale" F and G variances is a function of the correlation between F and G. Then the F test, $\sigma^2_{(F+G)}/(\sigma^2_F + \sigma^2_G)$ merely reflects the size and direction of the correlation between F and G.

Two common factors and no specific variances were conceived in the F and G data rather than one common factor and two specific variances, for if F and G covered the same content, with the obtained reliabilities, then they should have correlated $-.64$ instead of $-.20$. Either the lack of correlation between F and G was due to specific variances in each, or to acquiescence. The opposition among the items made it difficult to con-

sider the items as measuring different rather than opposite content.

Subjects and Regional Differences

As Messick and Jackson noted, the correlation of $-.20$ between F and G may have been unusual and by itself would suggest according to the corrected factor analysis that only about 25 per cent of the reliable variance of the F scale can be attributed to acquiescence. In reversed F scale studies carried out with other samples, and by other investigators including Messick and Jackson, correlations ranging from $-.42$ to $+.66$ have been found when a high negative relation consistently should have resulted between the F scale and a reversed F scale. Regional factors may be one source of variation in these correlations. Another may be different ways of reversing F scale items. Reversed items may assess different rather than opposing content.

In analyzing the relation between the F scale and a scale of social acquiescence, the author (2) found large regional differences. This work and the varying results of reversed F scale studies suggest that it is impossible to draw quantitative conclusions concerning authoritarianism and acquiescence without specifying the origins of the sample used as well as other conditions. Of three regions examined, Kansas, Maryland, and Louisiana, the lowest correlation between authoritarianism and acquiescence was found in the Louisiana sample (2). The correlations for Kansas, Maryland, and Louisiana were respectively: $.48$, $.49$ and $.16$. The means in F for the three samples

were: 3.92, 4.03, and 4.52. The means in social acquiescence followed the identical order: 28.7, 29.9 and 33.9. There was a correlation in means between groups as well as a correlation of scores within groups between acquiescence and authoritarianism. Since a Louisiana sample was employed in the study in question (1), the results as they now stand corrected may be more conservative than other investigators such as Jackson and Messick have obtained or will be likely to obtain with other samples.

Throw out the bathtub and the bath water, but save the baby.

REFERENCES

1. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.
2. BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296-299.
3. MESSICK, S. & JACKSON D. Authoritarianism or acquiescence in Bass' data. *J. abnorm. soc. Psychol.*, 1957, **54**, 424-426.
4. Symposium: Redefining authoritarianism. *Amer. Psychol.*, 1956, **11**, 365(a). (Abstract.)

Received November 26, 1956. Prior publication.



ERRATA

In the article "Rorschach and TAT indices of homosexuality in overt homosexuals, neurotics and normal males" by Anthony Davids, Mark Joelson, and Charles McArthur in the September 1956 issue of this Journal (*J. abnorm. soc. Psychol.*, 1956, **53**, 161-172), there is an error in the text. In the section on TAT results, under the heading of *Signs suggested for further confirmation* (p. 168), it states incorrectly (line 15) that the sign of a *strong unresolved attachment to a father or father figure* was scored in stories composed for Card 8. The sign was in fact scored in stories composed for TAT Card 7.

In the article entitled "Ethnic cleavage among young children" by W. E. Lambert and Y. Taguchi (this Journal 1956, **53**, 380-382) the following table should be substituted for Table 1 (p. 381), and one *p* value, for Picture-Taking Choices, has been corrected. The tests were one tailed since there was no apriori ground for expecting out-group preferences.

TABLE 1
TEST CHOICES OF OCCIDENTAL AND
ORIENTAL CHILDREN

	Candy-Giving Choices		Picture- Taking Choices		Picture- Choosing Choices	
	Occid.	Orient.	Occid.	Orient.	Occid.	Orient.
Occidentals	4	3	5	2	4	3
Orientals	0	6	0	6	0	6
Exact tests	$p = .048$		$p = .016$		$p = .048$	

The original tests of significance were calculated from the table in the form given above. Confusions of interpretation may arise from a reading of the original Table 1. The original table implies that the test is being made on the null hypothesis that in-group choice is independent of the ethnic background of the choosers. In reality, we were interested in determining whether there was a preponderance of in-group choices, that is, a tendency for persons to choose their own ethnic group.